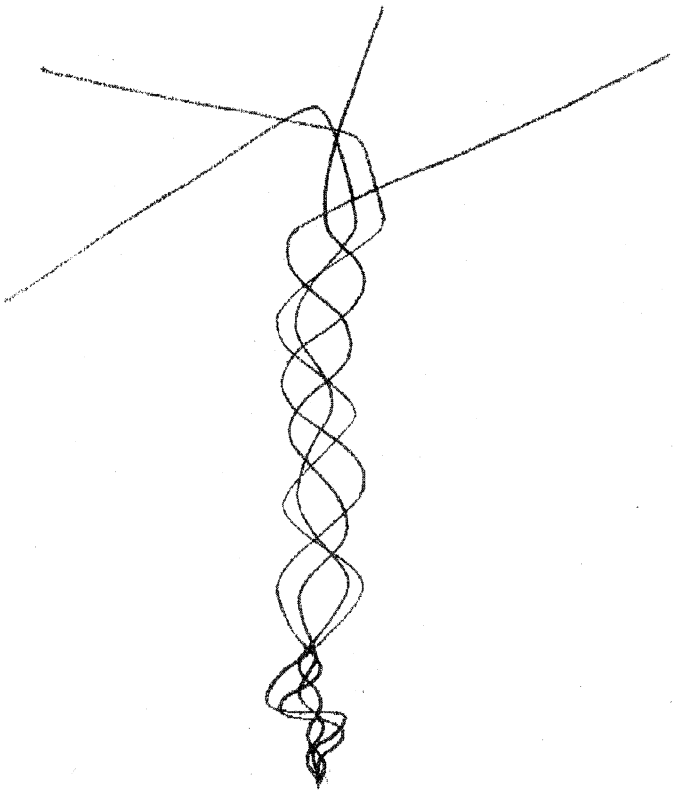
SIMULTANEOUS COMPONENT AND FACTOR ANALYSIS METHODS FOR TWO OR MORE GROUPS: A COMPARATIVE STUDY

Jan Niesing



DSWO PRESS

SIMULTANEOUS COMPONENT AND FACTOR ANALYSIS METHODS FOR TWO OR MORE GROUPS: A COMPARATIVE STUDY

M&T SERIES 30

Earlier publications in this series:

Jacqueline Meulman, Homogeneity Analysis of Incomplete Data.
M&T series 1, 1982
Pieter M. Kroonenberg, Three-mode Principle Component Analysis.
M&T series 2, 1983 (reprint 1989)
Jan de Leeuw, Canonical Analysis of Categorical Data.
M&T series, 1984
Ronald A. Visser, Analysis of Longitudinal Data in Behavioural and Social Research.
M&T series 4, 1985
John P. van de Geer, Introduction to Linear Multivariate Data Analysis.
M&T series 5, volume 1 & 2, 1986
Jacqueline Meulman, A Distance Approach to Nonlinear Multivariate Analysis.
M&T series 6, 1986
Jan de Leeuw, Willem Heiser, Jacqueline Meulman, Frank Critchley (editors), Multidimensional Data Analysis.
M&T series 7, 1987
Peter G.M. van der Heijden, Correspondence Analysis of Longitudinal Categorical Data.
M&T series 8, 1987
Jan van Rijckevorsel, The Application of Fuzzy Coding and Horseshoes in Multiple Correspondence Analysis.
M&T series 9, 1987
Abby Israëls, Eigenvalue Techniques for Qualitative Data.
M&T series 10, 1987
Eeke van der Burg, Nonlinear Canonical Correlation and Some Related Techniques.
M&T series 11, 1988
Kees van Montfort, Estimating in Structural Models with Non-normal Distributed Variables: Some Alternative Approaches.
M&T series 12, 1989
Jan T.A. Koster, Mathematical Aspects of Multiple Correspondence Analysis for Ordinal Variables.
M&T series 13, 1989
Catrien C.J.H. Bijleveld, Exploratory Linear Dynamic Systems Analysis.
M&T series 14, 1989
Henk A.L. Kiers, Three-way Methods for the Analysis of Qualitative and Quantitative Two-way Data.
M&T series 15, 1989
Stef van Buuren, Optimal Scaling of Time Series.
M&T series 16, 1990

Rien van der Leeden, Reduced Rank Regression with Structured Residuals.
M&T series 17, 1990
Rian van Blokland-Vogelesang, Unfolding and Group Consensus Ranking for Individual Preferences.
M&T series 18, 1991
Jacques J.F. Commandeur, Matching Configurations.
M&T series 19, 1991
Herbert Hoijtink, PARELLA, Measurement of Latent Traits by Proximity Items.
M&T series 20, 1991
Wendy J. Post, Nonparametric Unfolding Models, a Latent Trait Approach.
M&T series 21, 1992
Ivo A. van der Lans, Nonlinear Multivariate Analysis for Multiattribute Preference Data.
M&T series 22, 1992
Wim P. Krijnen, The Analysis of Three-Way Arrays by Constrained Parafac Methods.
M&T series 23, 1992
Marijtje A.J. van Duijn, Model for Repeated Counts
M&T series 24, 1993
Jos M.F. ten Berge, Least Squares Optimization in Multivariate Analysis.
M&T series 25, 1993
Patrick J.F. Groenen, The Majorization Approach to Multidimensional Scaling: Some Problems and Extensions.
M&T series 26, 1993
Peter Verboon, A Robust Approach to Nonlinear Multivariate Analysis.
M&T series 27, 1994
Monica Th. Markus, Bootstrap Confidence Regions in Non-Linear Multivariate Analysis.
M&T series 28, 1994
Frits Meijerink, A Nonlinear Structural Relations Model.
M&T series 29, 1995
Jan Niesing, Simultaneous Component and Factor Analysis Methods for Two or More Groups: A Comparative Study.
M&T series 30, 1997

Available form DSWO Press
Leiden University
P.O. Box 9555
2300 RB Leiden
The Netherlands
Tel. (071) 5273795 / 5273794
Fax (071) 5273619

SIMULTANEOUS COMPONENT AND FACTOR ANALYSIS METHODS FOR TWO OR MORE GROUPS: A COMPARATIVE STUDY

Jan Niesing

DEPARTMENT OF PSYCHOLOGIE
University of Groningen

1997 DSWO Press, Leiden University, The Netherlands

© Cover design: J. Niesing and Infofilm, Roger Busschots, Leiden
Printed by 'HAVEKA b.v.' and by
Printing office 'Karstens drukkers/designers', Leiden

CIP-Data Koninklijke Bibliotheek
Subject Headings: Principle Component Analysis, Factor Analysis, Cross Sectional Data Analysis
ISSN 0928-1088-30
ISBN 90-6695-134-6

ACKNOWLEDGEMENT

The past four years I have worked at my Ph.D. thesis with great pleasure. Many people have helped me with me and my study and the time for a grateful "thank you" has finally come.

First of all, I would like to thank Jos ten Berge and Henk Kiers for their intense coaching, their sharp and always useful comments, and making sure that this wild goose would keep focused on the study instead of flying around without direction (which is also fun, but does not get you anywhere). Furthermore they kept me from being too wordy, otherwise you would be holding only the first volume in a series of books about my research instead of the shortened version you are holding now.

For advices on the use of factor analysis methods and the analyses of the results from the experiments conducted I would like to thank Anne Boomsma and Tom Snijders, respectively. For reading the final manuscript I would like to thank the members of the review committee, prof. dr. G.J. Mellenbergh, prof. dr. W. Molenaar and prof. dr. W.J. Heiser.

A lot of pleasure came from the enjoyable atmosphere on the fourth floor of the Heymans building, where the department of Personality Psychology and the department of Statistics and Measurement Theory are located. Of the fourth floor colleagues I would especially like to mention my ex-room mates Francis Tuerlinckx, Jeroen Heres, and, last but not least, Gertrude Smit, with whom I shared my good and bad times.

As regards the graduation ceremony, I am very grateful to Wander Jager and Roy Bradwolff for assisting me and to get things organized.

Finally, I would like to thank my brother Wim and my parents Alie en Henk Niesing, for their strong and loving support throughout my study.

CONTENTS

CHAPTER 9 CONCLUSION AND DISCUSSION

CHAPTER 1

GENERALIZED COMPONENTS ANALYSIS

1.1 INTRODUCTION

In research in the social sciences, it is often the case that a number of variables (for instance a questionnaire) is measured on two or more different points in time, or at the same time in two or more groups of individuals (differing in some respect). For instance, one may have scores of children at six, eight and ten years of age on a set of intellectual tasks, or one may have scores of comparable groups of individuals from Zimbabwe and Sweden on a personality inventory. One of the first questions that comes to mind when analyzing data of these kinds is whether on the different time points or in the different groups the same factors are found. Are the same intellectual skills present at ages six, eight and ten? Are the same personality traits present in both Zimbabwe and Sweden?

An obvious way to answer these questions is to analyze the groups separately with a factor analysis technique followed by a Varimax rotation and compare the solutions found, using some similarity measure. Factor analysis techniques, such as Common Factor Analysis (CFA) and Principal Components Analysis (PCA) are useful methods to summarize the information, contained in the correlations between a large number of variables and in the scores on a large number of variables, respectively, in an efficient way. Because, in the present study, several generalizations from the method PCA are used, in Section 1.2, the method PCA will be presented.

Performing a Varimax rotation (Kaiser, 1958) on the pattern matrices, found with PCA on each group separately, and comparing the solutions (for instance using a congruence measure) has been the favorite approach of researchers in situations with two or more groups for many

years. The usefulness of this approach will be discussed in Section 1.3.

Another way of comparing the solutions, found with separate PCA's, is to rotate the two (or more) matrices to optimal agreement, using a Generalized Procrustes Rotation (GPR), see for instance Ten Berge (1977; 1986a; 1988) and Ten Berge and Knol (1984), and then applying a congruence measure. With a GPR, it can be investigated to what extent the same factors can be found in two or more groups. In Section 1.4, the GPR technique will be discussed. For both rotation techniques (Varimax and GPR), advantages and disadvantages will be summed up.

As will become clear, the method of using rotation techniques in the search for common factors suffers from a certain shortcoming. However, for the analysis of scores of two or more groups on the same variables, there are methods available that analyze the groups simultaneously, making the above approaches obsolete. A selection of these methods, used in the present study, will briefly be presented in Section 1.5 (for a full description, see Chapter 2), and in Section 1.6 the goal of the present study will be described.

In Section 1.7, reasons for ignoring some of the methods available for simultaneous analysis of two or more groups will be given, and in Section 1.8, one of the methods originally intended for inclusion in the present study will be discarded. In Section 1.9, the comparison of common factor analysis and component analysis for one group will be discussed. Finally, the approach taken in the present study will be presented in Section 1.10.

1.2 PRINCIPAL COMPONENTS ANALYSIS

When a researcher has scores of a group of n individuals on a set of m variables, Principal Components Analysis (PCA) is a popular technique for an exploratory analysis. In PCA, the variables are linearly combined (weighted sums of variables) in such a way that the amount of variance explained by these combinations is maximal. The combinations of variables are called components. In the present study, the term component will be

used for the methods making linear combinations of the variables, and the term factor will be used for the methods making use of hypothetical latent variables.

In the explanation to follow, standardized scores are used. Let the standard scores of n individuals on m variables be collected in an $n \times m$ matrix $\mathbf{Z}$ and the weights for combinations of the m variables to form q components in an $m \times q$ weights matrix $\mathbf{W}$. The component scores matrix $\mathbf{F}$ with the scores of the n individuals on the q components can now be written as $\mathbf{F}=\mathbf{ZW}$. PCA now minimizes the function

$$f(\mathbf{W},\mathbf{P}) = \| \mathbf{Z} - \mathbf{ZWP}' \|^2 = \| \mathbf{Z} - \mathbf{FP}' \|^2, \tag{1.1}$$

where the matrix $\mathbf{P}'$ provides the optimal regression weights for estimating matrix $\mathbf{Z}$ from $\mathbf{ZW}$.

The correlations between the variables are collected in the correlation matrix $\mathbf{R}=n^{-1}\mathbf{Z}'\mathbf{Z}$. Without loss of generality, the matrix $\mathbf{F}$ can be constrained to have standardized and uncorrelated columns. Subject to this constraint, the maximum of function f is found for $\mathbf{W}=\mathbf{K}_q\mathbf{\Lambda}_q^{-\frac{1}{2}}\mathbf{T}$ (cf. Ten Berge, 1993, p. 44), where the matrix $\mathbf{K}_q$ contains the first q columns of matrix $\mathbf{K}$, defined by the eigendecomposition $\mathbf{R}=\mathbf{K\Lambda K}'$, the matrix $\mathbf{\Lambda}_q$ is the upper left submatrix ($q \times q$) of matrix $\mathbf{\Lambda}$, and $\mathbf{T}$ is an arbitrary rotation matrix. The correlations between the variables and the components are collected in the $m \times q$ matrix $\mathbf{S}$, computed as $\mathbf{S}=n^{-1}\mathbf{Z}'\mathbf{F}=n^{-1}\mathbf{Z}'\mathbf{ZW}=\mathbf{RW}$; the correlations between the components are collected in $\mathbf{\Phi}$ ($q \times q$) and calculated as $\mathbf{\Phi}=n^{-1}\mathbf{F}'\mathbf{F}=n^{-1}\mathbf{W}'\mathbf{Z}'\mathbf{ZW}=\mathbf{W}'\mathbf{RW}=\mathbf{W}'\mathbf{S}=\mathbf{S}'\mathbf{W}$. The matrix $\mathbf{S}$ is called the structure matrix. Using the solution for $\mathbf{W}$ and the eigendecomposition of $\mathbf{R}$, any rotated PCA solution for the matrices $\mathbf{S}$ and $\mathbf{\Phi}$ can be written as $\mathbf{S}=\mathbf{K}_q\mathbf{\Lambda}_q^{\frac{1}{2}}\mathbf{T}$ and $\mathbf{\Phi}=\mathbf{T}'\mathbf{T}$ for some matrix $\mathbf{T}$ with $\mathrm{diag}(\mathbf{T}'\mathbf{T})=\mathbf{I}$.

The optimal solution for $\mathbf{P}$ is given as $\mathbf{P}=\mathbf{Z}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}=\mathbf{Z}'\mathbf{F}(n\mathbf{\Phi})^{-1}$ $=n^{-1}\mathbf{Z}'\mathbf{F}\mathbf{\Phi}^{-1}=\mathbf{S}\mathbf{\Phi}^{-1}$. The matrix $\mathbf{P}$ is called the pattern matrix. For any rotated PCA solution, the pattern matrix can be written as $\mathbf{P}=\mathbf{K}_q\mathbf{\Lambda}_q^{\frac{1}{2}}(\mathbf{T}')^{-1}$. From these descriptions, it can be easily seen that, for orthogonal components, the matrices $\mathbf{S}$ and $\mathbf{P}$ are equal. In this situation, this matrix is called the loadings matrix and denoted by $\mathbf{A}$, so $\mathbf{P}=\mathbf{S}=\mathbf{A}$.

From the component scores matrix $\mathbf{F}$ and the pattern matrix $\mathbf{P}$, found by PCA, the original variable scores matrix $\mathbf{Z}$ and correlation matrix $\mathbf{R}$

can be estimated as $\hat{\mathbf{Z}}=\mathbf{FP}'$ and $\hat{\mathbf{R}}=\mathbf{P\Phi P}'$, respectively. The estimate $\hat{\mathbf{Z}}$ is the best estimate of $\mathbf{Z}$, according to the least squares criterion. The fit of the solution can now be described as the amount of variance, explained by the components, found with PCA. This can be obtained as the sum of the variances of the columns of $\hat{\mathbf{Z}}$. Using the above expressions, the explained variance (EV) can be written as a function of matrices $\mathbf{P}$ and $\mathbf{\Phi}$, $\mathbf{S}$ and $\mathbf{\Phi}$ or $\mathbf{W}$ and $\mathbf{R}$, respectively, as

$$\text{EV} = n^{-1}\text{tr}(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}) = \text{tr}(\mathbf{P\Phi P}') = \text{tr}(\mathbf{S\Phi}^{-1}\mathbf{S}') = \text{tr}(\mathbf{W}'\mathbf{R}^2\mathbf{W}(\mathbf{W}'\mathbf{RW})^{-1}). \qquad (1.2)$$

Thus, when analyzing scores of a single group of individuals on a number of variables, PCA produces three matrices, each describing a relation between the scores on the variables and the scores on the components. These are: The components weights matrix $\mathbf{W}$ by which the components are defined as weighted sums of the variables; the structure matrix $\mathbf{S}$ with correlations between variables and components; and the pattern matrix $\mathbf{P}$ with regression weights for the reconstruction of the variables from the components. Each of the matrices $\mathbf{W}$, $\mathbf{S}$ and $\mathbf{P}$ can be used for interpretation of the components, although using the pattern matrix $\mathbf{P}$ is somewhat inconvenient. Because the pattern matrix conveys how the variables depend on the components rather than the other way around, it can only be used indirectly for interpretation of the components (cf. Brogden, 1969; Gorsuch, 1983, p. 207, see Ten Berge, 1986a). Furthermore, Ten Berge (1986a) has shown that the columns of $\mathbf{P}$ are not uniquely defined for each separate component. That is, rotations which leave the position of one component unaltered do affect the column of the matrix $\mathbf{P}$ which corresponds to that component. Still, the pattern matrix is widely used for interpretation of components.

The solutions, found with PCA, are typically rotated to some form of simple pattern, before comparisons over groups are made. The favorite rotation for many years has been the Varimax rotation. The Varimax rotation technique was developed in order to facilitate the interpretation of orthogonal components. Kaiser (1958) suggested maximizing the variance of the squared loadings summed across components, in order to simplify the loadings matrix $\mathbf{A}$ ($=\mathbf{P}=\mathbf{S}$). Specifically, that rotated matrix $\mathbf{AT}$ is sought for which the sum (over columns) of

columnwise variances of squared loadings is maximized. Its goal is to rotate the components in such a way that each component contains large loadings for some variables and small loadings for the other variables. The Varimax rotated loadings matrix thus contains as many near zero and large elements as possible, and as few as possible elements in between. Applying this rotation, therefore, generally increases the interpretability of the resulting components, compared to the unrotated components.

In presenting PCA, so far only the situation when one has only a single group of scores has been discussed. What happens when the scores of two or more groups of respondents on the same variables need to be analyzed? The most straightforward approach would be to perform the procedure presented above on each group separately and see whether the same components are found. Thus, a PCA would be performed on each group separately (called PCA–sep), followed by a Varimax rotation on each of the solutions. The usefulness of this procedure is examined in the next Section.

1.3 COMPARING VARIMAX ROTATED COMPONENTS

In order to compare the interpretations of corresponding components in the different groups, Tucker's congruence coefficient (Tucker, 1951) can be calculated for the corresponding columns of the loading matrices. This is a measure for the proportionality of columns. When the congruence of the components found with the two separate analyses is higher than .85, one can conclude that the components will get the same interpretations in both groups (see Haven and Ten Berge, 1977). However, when the congruence of the components found with the two analyses is lower than .85, no conclusion can be drawn. Although a congruence lower than .85 does imply the components *found* in the groups are essentially different, it is still possible that the same components are indeed present in both groups. For instance, a different rotation technique might result in components with a higher congruence.

1.4 APPLYING THE GENERALIZED PROCRUSTES ROTATION

To see if indeed the same components are present in different solutions, Generalized Procrustes Rotation (GPR) can be used. GPR rotates p matrices ($p \geq 2$) toward a best least-squares fit. Specifically, if $\mathbf{A}_1,...,\mathbf{A}_p$ is a set of p matrices of order $m \times q$ ($m \geq q$), GPR finds orthonormal matrices $\mathbf{T}_k$ ($k=1,...,p$) for which the function

$$f(\mathbf{T}_1,...,\mathbf{T}_p) = \sum_{k=1}^{p} \sum_{l=1}^{p} \| \mathbf{A}_k\mathbf{T}_k - \mathbf{A}_l\mathbf{T}_l \|^2 \tag{1.3}$$

is minimized.

If after GPR all congruence coefficients between corresponding columns of $\mathbf{A}_k\mathbf{T}_k$ and $\mathbf{A}_l\mathbf{T}_l$ are high, it can be concluded that the same components are present in the solutions. However, if a GPR fails to yield (nearly) equivalent components, it is still possible that common components do exist. This can be argued in two ways. Firstly, all rotational techniques are, in their search for common components, limited to already found component solutions. It is conceivable that, when a larger number of components is taken, the same components suddenly *do* appear in both groups, because they were merely ordered differently (in terms of the amount of variance they explain) in both groups. Secondly, GPR can only use orthogonal rotations. It is possible that, by using oblique rotations of the solutions, one does find the same components. However, allowing for totally unconstrained oblique rotations in GPR would lead to degenerate solutions where all columns of a matrix are transformed into the same vector. Hence, as soon as one is willing to employ totally unconstrained oblique transformations, it no longer makes sense to use GPR.

A more general problem with GPR, and this again goes for all rotational methods, is that when one is comparing components found, some measure for comparison has to be used. Mostly this will be the congruence criterion (for a description of the congruence criterion, see Section 3.4.1). The problem that is always present, when using such a measure, is for what values of the congruence criterion two components must be judged equal or unequal. The threshold value that has been advocated for the congruence criterion by Haven and Ten Berge (1977) is .85, but what does

one do when the congruence value found falls just below this value, say .84? A researcher, facing this value, could very well be inclined to judge the components as being "almost" equal. Hence, with the use of this measure, no strong conclusions can be drawn.

It was mentioned that when, after GPR, congruences are high, it can be concluded that the same components are present in the solution. However, because GPR attempts to rotate matrices towards one another, some caution is advised. Horn and Knapp (1973) have found that, with *random* target matrices, GPR succeeded in rotating solutions from a factor analysis to such a position that 84% of the hypothesized 'salient' (but randomly chosen) loadings were recaptured from the solution. Thus, GPR might indicate that common components exist in two groups, when in fact there are no (strong) common components.

For these reasons, the GPR does not seem to be the ideal approach. Thus, a different approach seems necessary, that, for one thing, also looks outside the range of rotations of the factor solutions. Two types of techniques for this will be presented in the next section.

1.5 METHODS FOR SIMULTANEOUS ANALYSIS IN TWO OR MORE GROUPS, USED IN THE PRESENT RESEARCH

In the methods presented above, first a q–dimensional solution was derived for each group separately, using a PCA on each group separately, followed by some rotational method, and secondly these solutions were compared. A way to avoid the problems accompanying this approach, is to simultaneously analyze the groups, with a method of analysis that puts an equality constraint on one of the matrices that can be used for interpretation of the components (**W**, **P** and **S**). With such a method it is no longer necessary to decide whether or not components with equal interpretations can be found in the groups, because equality is enforced by the method of analysis. For each of the methods coming up, one of the matrices **W**, **P** and **S** is used for the interpretation of the components.

To see how easy it is to find equal components, let the component

interpretation for a group be based on the weights matrix, found with PCA in one group. To find the same components in a different group, one can simply use this weights matrix for combining the variables in this other group (provided the same variables are used in both groups), thus by definition creating components with exactly the same interpretation in both groups. So the question whether or not it is possible to find the same components across groups is a trivial one: It is always possible to obtain the same components in all groups, by simply defining them in the same way in all groups (see Ten Berge, 1986b), although they need not account for much variance in all groups. A more meaningful question is thus how much of the variance is explained by such components in each group. This is directly related to the aim of PCA: The aim of PCA should *not* be thought of as to find *the* components present in a particular data set (because every conceivable component is present in a data set), but to find a (small) number of components that represent the data as well as possible, by accounting for the maximum amount of variance in the data. We are now in a position to introduce the methods for simultaneous analysis of two or more groups, used in the present study.

Firstly, there are three types of Simultaneous Components Analysis (SCA) methods. With these methods, correlated components are found that explain a maximum amount of variance over the groups of individuals, under the constraint that either the components weights matrix, the pattern matrix, or the structure matrix is kept equal (or at least columnwise proportional) over all groups (see Levin, 1966; Millsap and Meredith, 1988; and Kiers and Ten Berge, 1989; 1994a, respectively), thus ensuring that components with the same interpretation (based on a common weights matrix **W**, a common pattern matrix **P** or a common structure matrix **S**, respectively) are found. These methods are called SCA–W, SCA–P and SCA–S, respectively.

Secondly, there are methods that impose common factors in statistical models for covariance or correlation matrices. These models are fitted using the maximum likelihood principle and will be called ML–methods. In the present study, one ML–method was investigated, namely Simultaneous Factor Analysis in Several Populations (SIFASP–ML, Jöreskog,

1971). Another ML–method, called Partial Common Principal Components Analysis (PCPC, Flury, 1987), originally intended for inclusion in the present study, was discarded from the study in an early stage. This method and reasons for discarding it are presented in Section 1.7.

The methods mentioned above are not the only available methods. In cases where a researcher wants to analyze scores of two or more groups of individuals on the same variables, several other methods for simultaneous factor analysis methods can be used. This study is limited to the four methods, presented above. These methods have been selected because of their promising features. For a detailed description of the methods see Chapter 2.

1.6 GOALS OF THE PRESENT STUDY

When analyzing scores from different groups of respondents on the same variables by a simultaneous components or factor analysis, researchers are faced with a variety of available methods, as we have seen above. Which method is chosen by a researcher seems to be a matter of taste. A choice between the methods on scientific grounds appears to be impossible, because there has been almost no research, systematically comparing these methods. In an attempt to fill this void, in the present study, the usefulness of a selection of the available methods of analysis is examined in a 'comparative commodities research', by applying these methods on constructed data of which the factor solutions are known. In this study, these known solutions are defined as the desired outcomes for the four methods. The main question in the present study is: "To what extent are the methods of analysis doing what they are supposed to do, under different circumstances? The circumstances are manipulated through simulation. The goal of this comparative study is therefore to verify which method is to be preferred under what circumstances and to come up with directions for selecting an appropriate method of simultaneous factor analysis, depending on the circumstances. It is, for instance, conceivable that one (or several) of the methods outperforms some of the

others when the number of variables is small or when the factors are only faintly present in the data. On the other hand, some of the methods can be especially suited for detecting differences between populations while the other methods are not.

The above methods are not the only methods available for the analysis of scores of two or more groups on a set of variables. A number of alternative methods, not dealt with in the present study, will be described in the next section, along with the reasons for discarding them.

1.7 OTHER METHODS FOR THE SIMULTANEOUS ANALYSIS IN TWO OR MORE GROUPS: IDIOSCAL, PARAFAC2 AND INDSCAL

Besides the techniques for simultaneous components or factor analysis, presented in Section 1.5, several other techniques, that might be considered for inclusion in the present study, have been suggested over the last decades. These techniques, called IDIOSCAL, PARAFAC2 and INDSCAL, respectively, will now be presented in short. After that, it will be discussed why these techniques were not included in the present study.

The three methods, presented next, can be viewed as three-way Multidimensional scaling (MDS) methods. Therefore, in the next section, the theory behind MDS will be briefly introduced. At this point, it must be noted that MDS is a technique usually applied to matrices (of order $n \times n$, where n is the number of individuals) containing scores of observer k ($k=1,\ldots,p$), and describing the (dis)similarity between the n individuals, based on the scores of each person on m variables. Because the methods presented in Section 1.5 all focus on the analysis of matrices containing covariances or correlations between m variables, obtained for p ($p \geq 2$) groups of individuals, MDS and the three methods of analysis for MDS will be presented in a form directly applicable to this kind of data.

1.7.1 Multidimensional Scaling

Many situations arise in the social and behavioral sciences in which several covariance matrices for the same variables are available for two or more groups. One would generally like to account for all of these data matrices in a single comprehensive analysis, based on a plausible model. Multidimensional scaling (MDS) is a useful mathematical tool for this, because it can use covariance measures as input, does not require a priori knowledge of the attributes of the variables to be scaled and provides a space that reveals dimensions (components) relevant to the groups. That is, MDS procedures represent variables judged experimentally similar to one another as points close to each other in a resultant spatial map.

As said, the most typical application of MDS, in the present context, arises when covariance matrices are available from two or more groups of individuals. However, instead of groups of individuals, the different matrices may also be associated with other types of data sources, such as different experimental conditions, tasks, or occasions. The term 'three–way MDS' is used to characterize the models and methods, used in these situations. The input for any three–way method consists of a three–way array $\mathbf{C}$ that can be thought of as comprising a set of p (≥ 2) $m \times m$ symmetric slices $\mathbf{C}_1,...,\mathbf{C}_p$, for instance, covariance matrices.

The three–way MDS method IDIOSCAL is presented in Section 1.7.2. The three–way MDS methods PARAFAC2 and INDSCAL, presented in Section 1.7.3 and 1.7.4, respectively, can be considered as models imposing restrictions on the model IDIOSCAL. The three models presented can be viewed as differing essentially in how they decompose the covariances for group k, $k=1,...,p$.

1.7.2 IDIOSCAL

The most general model in this class of three–way MDS models has been called (by Carroll & Chang, 1972) the IDIOSCAL (*I*ndividual *DI*fferences *I*n *O*rientation *SCAL*ing) model. In matrix notation, this model

can be expressed as

$$\mathbf{C}_k = \mathbf{A}\mathbf{H}_k\mathbf{A}' + \mathbf{E}_k, \tag{1.4}$$

$k=1,...,p$, where $\mathbf{H}_k$ is a $q \times q$ symmetric positive semi-definite (p.s.d) matrix, $\mathbf{A}$ is an arbitrary $m \times q$ $(m \geq q)$ matrix, and $\mathbf{E}_k$ is a matrix containing the part not explained by the model. IDIOSCAL minimizes the function

$$\sigma_1(\mathbf{A},\mathbf{H}_1,...,\mathbf{H}_p) = \sum_{k=1}^{p} \| \mathbf{C}_k - \mathbf{A}\mathbf{H}_k\mathbf{A}' \|^2. \tag{1.5}$$

Kiers, Cléroux and Ten Berge (1994) proposed an algorithm that minimizes the IDIOSCAL function and converges monotonically, for p.s.d. matrices $\mathbf{C}_k$, $k=1,...,p$.

Ten Berge, Bekker and Kiers (1994) considered the TUCKALS2 (T2) algorithm applied to the IDIOSCAL problem, as first proposed by Kroonenberg and De Leeuw (1980). Ten Berge et al. proved that, in case TUCKALS2 is applied to p.s.d. matrices, the TUCKALS2 algorithm does converge to a proper IDIOSCAL solution.

Note that the solution for the matrix $\mathbf{A}$ for the IDIOSCAL problem is not unique, although all solutions span the same column spaces.

1.7.3 PARAFAC2

In order to introduce PARAFAC2 properly, first the method from which it was derived, PARAFAC1 (*PARA*llel *FAC*tors), will be presented. PARAFAC1 is an interesting generalization of principal components analysis for the situation where the same variables have been measured on the same observation units (individuals, groups) a number of times (Harshman, 1970; Harshman & Lundy, 1984). One of the main features of PARAFAC1 is that it gives factors that (under mild conditions) are unique up to scaling and permutation. However, the standard PARAFAC1 method cannot be used for analyzing data from different groups. For an exploratory analysis of such cross-sectional data, taking into account that the same variables were measured in all groups, Harshman (1972) proposed to fit the PARAFAC2 model to the covariance matrices obtained in the different groups (see also Carroll & Wish, 1974, p. 94–96; Harshman & Lundy, 1984, p. 187).

If $\mathbf{C}_k$ denotes the $m \times m$ covariance matrix at occasion k, then the PARAFAC2 model can be described as

$$\mathbf{C}_k = \mathbf{A}\mathbf{D}_k\mathbf{H}\mathbf{D}_k\mathbf{A}' + \mathbf{E}_k, \tag{1.6}$$

$k=1,...,p$, where $\mathbf{A}$ denotes an $m \times q$ matrix of 'loadings' for the m variables on a set of q 'latent factors', $\mathbf{D}_k$ denotes a $q \times q$ diagonal matrix containing 'saliences' for the different factors for $\mathbf{C}_k$, $\mathbf{H}$ denotes a $q \times q$ symmetric (and usually p.s.d) matrix with relations between the factors, and $\mathbf{E}_k$ is a matrix with error terms. For details on the interpretation of the various parameter sets, the reader is referred to Harshman (1972), Carroll and Wish (1974, pp. 94–96), Dunn and Harshman (1982), or Kroonenberg (1983, pp. 53–56), among others.

Compared with the IDIOSCAL model, Harshman's (1972) PARAFAC2 model is the specific case of IDIOSCAL in which the matrix $\mathbf{H}_k$ is constrained to be of the form

$$\mathbf{H}_k = \mathbf{D}_k\mathbf{H}\mathbf{D}_k, \tag{1.7}$$

where $\mathbf{D}_k$ is a diagonal matrix.

The PARAFAC2 model is fitted to a set of covariance matrices in the least squares sense by minimizing

$$\sigma_2(\mathbf{A},\mathbf{H},\mathbf{D}_1,...,\mathbf{D}_p) = \sum_{k=1}^{p} \| \mathbf{C}_k - \mathbf{A}\mathbf{D}_k\mathbf{H}\mathbf{D}_k\mathbf{A}' \|^2. \tag{1.8}$$

Kiers (1993) proposed an alternating least squares algorithm for PARAFAC2, using an adjusted version of the three-way DEDICOM (DEDICOM3) algorithm, alternately updating the matrices $\mathbf{H}$, $\mathbf{A}$, and $\mathbf{D}_1,...,\mathbf{D}_p$. Kiers found the PARAFAC2 algorithm to show good convergence properties, always explaining at least 99% of variance in errorless data constructed with $\mathbf{H}$ p.s.d. and according to (1.6), although the procedure converged very slowly when the function value in (1.8) approached zero. Kiers also demonstrated the well known nonuniqueness of the PARAFAC2 solutions in case only two covariance matrices are analyzed.

Ten Berge and Kiers (1996) have shown that PARAFAC2 is usually unique with four matrices, in the rank two case. With three matrices, the PARAFAC2 solution is unique when an additional assumption is introduced. If, for instance, the diagonal matrices $\mathbf{D}_k$, $k=1,2,3$, are constrained to

be nonnegative, three matrices are enough to have uniqueness in the rank two case of PARAFAC2. See also Harshman and Lundy (1996) for uniqueness proofs of PARAFAC2 for ranks larger than 2.

1.7.4 INDSCAL

INDSCAL (*IN*dividual *D*ifferences *SCAL*ing) is probably the most widely used three–way MDS method. INDSCAL decomposes the covariance matrix $\mathbf{C}_k$ as

$$\mathbf{C}_k = \mathbf{A}\mathbf{D}_k\mathbf{A}' + \mathbf{E}_k, \tag{1.9}$$

k=1,...,p, with $\mathbf{D}_1,...,\mathbf{D}_p$ diagonal and nonnegative (Carroll & Chang, 1970, p. 284; Carroll, 1972, p. 111; Ten Berge, 1993, p. 64). In the present situation, INDSCAL is the special case of IDIOSCAL in which the matrix $\mathbf{H}_k$ is diagonal and nonnegative.

The INDSCAL model is based on the premise that for all groups the same matrix $\mathbf{A}$ with loadings of the variables on q dimensions is used, but that for each group the idiosyncratic weights attached to these dimensions differ. These weights are the diagonal elements of $\mathbf{D}_1,...,\mathbf{D}_p$.

The problem how to fit the INDSCAL model in the least squares sense amounts to minimizing

$$\sigma_3(\mathbf{A},\mathbf{D}_1,...,\mathbf{D}_p) = \sum_{k=1}^{p} \| \mathbf{C}_k - \mathbf{A}\mathbf{D}_k\mathbf{A}' \|^2. \tag{1.10}$$

This problem has not been solved directly. The most popular approach, due to Carroll and Chang (1970), is based on a technique called splitting (De Leeuw & Heiser, 1982). That is, the two appearances of $\mathbf{A}$ in (1.10) are represented by different matrices, $\mathbf{A}$ and $\mathbf{B}$. These are optimized independently, along with $\mathbf{D}_1,...,\mathbf{D}_p$, by means of the "CANDECOMP" algorithm which minimizes $\sum_{k=1}^{p} \| \mathbf{C}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}' \|^2$. After convergence of CANDECOMP, it is hoped that the $\mathbf{A}$ and $\mathbf{B}$ obtained happen to be equal. In practice, this always appears to be the case. However, Ten Berge and Kiers (1991) have shown that essentially different matrices $\mathbf{A}$ and $\mathbf{B}$ can be found at stationary points of the CANDECOMP function, when Gramian matrices $\mathbf{C}_k$ are analyzed. They did not arrive at a proof for the equality of $\mathbf{A}$ and $\mathbf{B}$ at the global minimum of the CANDECOMP function, nor did they encounter any counterexamples.

Also, one must hope that the solutions obtained for $\mathbf{D}_1,...,\mathbf{D}_p$ have nonnegative diagonal elements throughout. This is not always the case in practice. It is possible to avoid negative saliences by imposing nonnegativity constraints (for instance see Carroll, Pruzansky & Kruskal, 1980; Ten Berge, Kiers & Krijnen, 1993).

A special property of INDSCAL is that its solutions are unique up to mild conditions (Harshman, 1972; see also Ten Berge & Kiers, 1991).

1.7.5 Reasons for not selecting the MDS methods

The three methods, presented above, were discarded from inclusion in the present study on account of practical limitations. The methods do not minimize the residual variance of the data in one way or another. Instead of presenting a model for the data itself, covariance matrices (or correlation matrices) are fitted. This kind of indirect fitting does not tie in nicely with SCA.

1.8 DISCARDING THE PARTIAL COMMON PRINCIPAL COMPONENTS METHOD

As said, the ML–method, called Partial Common Principal Components Analysis (PCPC, Flury, 1987), originally intended for inclusion in the present study, was discarded from the study in an early stage. Next, a short overview of this method will be given, together with the reasons for discarding it.

1.8.1 Partial Common Principal Components (PCPC)

Flury (1984) conceived of PCA as a method of diagonalizing a covariance matrix. Accordingly, he has developed a generalization of PCA, aimed at diagonalizing a set of $m \times m$ covariance matrices $\mathbf{\Sigma}_k$ (using his notation) of p samples simultaneously, called the Common Principal Components model (CPC). In CPC, it is assumed that the p population

covariance matrices have the same m eigenvectors, that is, the same orthonormal matrix diagonalizes all matrices $\mathbf{\Sigma}_k$ simultaneously. For the CPC model, Flury derived a normal theory maximum likelihood estimator.

As a modification of the CPC model, Flury (1987) assumed that only q out of m eigenvectors are common to all matrices $\mathbf{\Sigma}_k$, while the remaining $m-q$ eigenvectors are specific to each group. This is called the Partial Common Principal Components model (PCPC). Flury (1987) defined the PCPC model as

$$\mathbf{\Sigma}_k = \mathbf{B}^{(k)}\mathbf{\Delta}_k\mathbf{B}^{(k)'} \quad (k=1,...,p), \tag{1.11}$$

with $\mathbf{B}^{(k)} = (\mathbf{B}_1:\mathbf{B}_2^{(k)})$, and $\mathbf{\Delta}_k$ diagonal, where all $\mathbf{B}^{(k)}$ are orthonormal $m\times m$ matrices, the common part $\mathbf{B}_1$ has dimension $m\times q$ ($q \leq m-2$), and $\mathbf{B}_2^{(k)}$ has dimension $m\times(m-q)$, $k=1,...,p$. It is the common part $\mathbf{B}_1$ that is currently of interest, because the columns of $\mathbf{B}_1$ define the factors common for all groups.

For the PCPC model, the common likelihood function is given as

$$L(\mathbf{\Sigma}_1,...,\mathbf{\Sigma}_p) = c \times \Pi_{k=1}^{p} \exp(\mathrm{tr}(-\tfrac{1}{2}N_k\mathbf{\Sigma}_k^{-1}\mathbf{C}_k))|\mathbf{\Sigma}_k|^{-\frac{1}{2}N_k}, \tag{1.12}$$

in which c is a constant and N_k is n_k-1. Equivalent to maximizing the likelihood function is the minimization of

$$g(\mathbf{\Sigma}_1,...,\mathbf{\Sigma}_p) = -2 \log L(\mathbf{\Sigma}_1,...,\mathbf{\Sigma}_p) + 2 \log c =$$

$$\sum_{k=1}^{p} N_k(\log|\mathbf{\Sigma}_k|+\mathrm{tr}\mathbf{\Sigma}_k^{-1}\mathbf{C}_k). \tag{1.13}$$

1.8.2 Reasons for discarding PCPC

Flury (1987) states that "Solving the likelihood equations for PCPC is extremely laborious and no simple or elegant method has yet been found to accomplish this. [...] However, an approximate solution can easily be obtained, provided that maximum likelihood estimates for the ordinary common principal component model are available. The approximation is based on the observation that, if the partial model holds, the q common components are estimated almost correctly in the ordinary common principal component model, irrespective of the other components that may

be specific in each population (p. 61)." This means that, because in the present study, the point of interest lies in the q common components, the solution found with PCPC is nothing else than a selection of q components from the solution, found with the CPC model. Which q components must be chosen from the m–dimensional solution of the CPC model, however, is not univocally clear.

Two of the possibilities, suggested by Flury (personal communication, 1994), are (1) one can have some idea about a component one expects to find: (2) one can look at the correlations between the components found in the different groups. When two components correlate high in one group and not at all in another, you can not consider these components to be common for both groups and therefore both must be abandoned. However, these guidelines depend to a large extent on subjective decisions.

Ten Berge (1995) has criticized the criterion of diagonalization, used in PCPC (see also Flury, 1988, pp. 205–210). Ten Berge states that "It is by no means granted that components which (partially) diagonalize the covariance or correlation matrices reflect the most important sources of variation in the data." In a small example a clear demonstration of this is given. Let

$$\mathbf{R}_1 = \begin{pmatrix} 1 & .8 & .6 & 0 \\ .8 & 1 & .7 & 0 \\ .6 & .7 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{R}_2 = \begin{pmatrix} 1 & .6 & .75 & 0 \\ .6 & 1 & .85 & 0 \\ .75 & .85 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and suppose a solution with q=1 is desired. The PCPC model attains the best diagonalizing solution when $\mathbf{B}$=(−.0002 −.0002 .0004 1.0000)'. Clearly, the resulting common component is connected to the fourth variable, rather than with the first three variables. Thus, it fails to reflect the major source of variation underlying the data. This demonstrates a serious drawback of the criterion of diagonalizability.

For the reasons cited above it was decided to discard PCPC from the present study.

1.9 FACTOR ANALYSIS VERSUS COMPONENTS ANALYSIS IN THE ONE GROUP SITUATION

In the present study, generalizations of both components analysis and factor analysis are studied. Factor analysis and principal components analysis use different models for a rather similar analysis of the relations among a number of variables, both describing these relations by means of loadings of the variables on mediating new variables (factors and components, respectively), as well as by correlations with and among these new variables. Considering the common purpose, many researchers have addressed the issue whether one should use a components analysis or a factor analysis, when analyzing the data for one group. Unfortunately, the issue has not been solved. In the present section, some results of the discussion and some of the findings of research concerning factor and component analyses of a single group of scores will be summarized, mainly on the basis of Velicer and Jackson's (1990) review article. It was stated by Velicer and Jackson (1990) that the choice between the methods was not obvious because they serve a similar purpose and share many characteristics, but it can also be said that the choice between the methods is not obvious because they serve different purposes and differ on certain characteristics (Widaman, 1993, p.308).

Velicer and Jackson (1990) discussed a number of theoretical and practical issues. The practical issues included: the degree of numeric similarity between solutions from the two methods, some common rules for the number of factors to be retained, problems with improper solutions (only for factor analysis), and comparisons in the computational efficiency. The theoretical issues included: the factor indeterminacy issue, the differences between exploratory and confirmatory procedures, and the issue of latent versus manifest variables.

With respect to the similarity issue, they state that "when the same number of components or factors are extracted, the results from different types of component or factor analysis typically yield highly similar results. Discrepancies are rarely, if ever, of any practical importance in subsequent interpretations. [...] Arguments that the slightly larger

loadings are *better* or *worse* are not likely to convince anyone."

In their discussion of some common rules for the number of factors to be retained, Velicer and Jackson's final conclusion is that "alternative, more accurate procedures, should replace the common, but problematic procedures."

In factor analysis, it is frequently the case that solutions are improper (for instance, because uniqueness estimates are zero or negative, making the solution, which they are part of, uninterpretable). This is often the case in so-called boundary conditions. These are conditions, such as the presence of only low saturation variables, small sample size, and low variable to factor ratio. The occurrence of improper solutions is usually treated as an annoyance, but according to Velicer and Jackson, could turn into an advantage for factor analysis compared to component analysis, when treated as a diagnostic, signaling that the present model is incorrect for the presented data set.

Concerning the computational issue, Velicer and Jackson state that "in general, component analysis will involve less computing processor time" and "in addition, can be performed on larger variable sets."

Coming to the theoretical issues next, Velicer and Jackson address the factor indeterminacy problem: "Factor indeterminacy is the inability to determine uniquely the common and unique factor variables of the common factor model from the uniquely defined observed variables because the number of observed variables is smaller than the number of common and unique factors" (Mulaik & McDonald, 1978, p.177). Extensive reviews of this problem are provided by Steiger (1979), Steiger and Schönemann (1978) and McDonald and Mulaik (1979). The implications of the factor indeterminacy issue are still being debated, although most writers simply ignore the issue.

The latent versus manifest variables issue concerns the fact that component analysis, being a manifest variable procedure that makes weighted linear composites of the observed variables, is limited to the space of the observed variables, while factor analysis, being a latent variable procedure, is not. Latent variables are estimates of unobserved, underlying, error free variables that (with the addition of sampling

error) account for the observed variables. The major assumed advantage of a latent variable model is the perceived added degree of generalization to other, unsampled variables. However, the existing empirical evidence does not support this claim. In fact, Velicer and Jackson (1990, p. 19) report that the slight differences found favor component analysis.

Some authors, according to Velicer and Jackson, "have made the distinction between component analysis as an exploratory procedure and factor analysis as a confirmatory procedure. The basis of this distinction is unclear, because almost identical operations are employed when either technique is employed in an exploratory manner." They prefer exploratory analytic approaches except for those cases where a well defined theory exists or where there are two (or more) competing legitimate theories, because exploratory approaches avoid confirmation bias, do not force a theory–oriented approach prematurely, and represent a conservative strategy.

The major conclusion of the Velicer and Jackson article is that there is little basis to prefer either component or factor analysis, because although there are some differences, for practical purposes, the choice of method will not greatly affect empirical results or substantive conclusions.

Widaman (1993) compared common factor analysis and principal component analysis, using error–free population data to ensure that sampling variability would neither obscure results nor be available as an alternative explanation for trends noted. He found, among other things, when using component analysis, nonzero pattern loadings were larger than the population loadings and the estimates of component intercorrelations were consistently lower than in the population. Common factor analysis, however, reproduced pattern loadings and interfactor correlations accurately. Widaman further showed that component analysis and common factor analysis define different parameters in the population, and therefore estimate different parameters in samples from that population, as was also shown by Snook and Gorsuch (1989).

Because, in the one group situation, there is no agreement on which method is the best, we have no expectations about the relative

performance of the generalizations of these methods to the two or more groups situation. Therefore, no expectation as to which is the best method for the two or more groups situation is given.

1.10 THE ORGANIZATION OF THE PRESENT STUDY

In Chapter 2, the four methods under investigation (SCA-W, SCA-P, SCA-S and SIFASP-ML) will be presented. For each method, the rotational methods used for simplification of the solution will be presented and its choice motivated. At the end of Chapter 2, some comparisons of the methods on theoretical grounds will be discussed.

In Chapter 3, the method of data construction used in the present study, will be presented. Two data categories will be defined, one for samples from one population and one for samples from two (or more) populations, and a measure for quantifying overall differences between populations will be presented.

In Chapter 4, the manipulations of the independent variables in this study will be described shortly, and measures used for dimension indication for the SCA-methods and SIFASP-ML will be described. Next, the success criteria, used in the present study, are introduced. At the end of Chapter 4, two pilot experiments will be presented, in which the necessary and useful range of amount of error to be added to scores is explored and a suitable convergence criterion for iterative procedures, used in two of the methods under investigation (SCA-W and SCA-S), is determined.

In Chapter 5, the main experiment comparing the three SCA-methods is presented. Findings in this experiment were used as a basis for comparison in following experiments.

In Chapter 6, the SCA-methods are being compared with the method SIFASP-ML. Findings in this experiment were used as a basis for comparison in following experiments.

In Chapter 7, the SCA-methods are compared in an experiment simulating a longitudinal process, in which some components grow weaker

over time, while others grow stronger.

In Chapter 8, the performance of the SCA–methods and the method SIFASP–ML is investigated in three special conditions: analyzing four groups simultaneously, large samples, and overlapping factors.

In Chapter 9, final conclusions are drawn and some guidelines are presented for how to use the most successful method(s) from the present study in practical research situations.

CHAPTER 2

THE METHODS UNDER INVESTIGATION

2.1 INTRODUCTION

Two types of methods, used for the simultaneous analysis of two or more groups (three SCA–methods and one ML–method), will be presented separately, starting with the SCA–methods. To ease the presentation, the following definitions are useful:

p = number of groups,
n_k = number of individuals in group k, with $N = \sum_{k=1}^{p} n_k$,
m = number of variables,
q = number of components ($q<m$),
$\mathbf{X}_k$ = $n_k \times m$ matrix with deviation scores in group k,
$\mathbf{Z}_k$ = $n_k \times m$ matrix with standard scores in group k,
$\mathbf{C}_k$ = covariance matrix of $\mathbf{X}_k$,
$\mathbf{R}_k$ = correlation matrix of $\mathbf{X}_k$, $\mathbf{R}_k = n_k^{-1}\mathbf{Z}_k'\mathbf{Z}_k$,
$\mathbf{W}_k$ = weights matrix ($m \times q$) for group k,
$\mathbf{P}_k$ = pattern matrix ($m \times q$) for group k,
$\mathbf{S}_k$ = structure matrix ($m \times q$) for group k.

For the factor analysis methods, additional definitions will be presented separately to avoid confusion.

The SCA–methods will be introduced using standard scores, while the ML–method will be introduced using deviation scores. This is done because this is the way the methods are usually presented.

2.2 SIMULTANEOUS COMPONENTS ANALYSIS METHODS

Because the SCA–methods are all generalizations of PCA, first a joint formulation of PCA, when performed on p groups separately, will be

given. When one has two or more populations, separate PCA's can be performed on each group, by minimizing the function

$$f_{sep}(\mathbf{W}_1,...,\mathbf{W}_p,\mathbf{P}_1,...,\mathbf{P}_p) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{W}_k\mathbf{P}_k' \|^2, \quad (2.1)$$

for a given q ($<m$). This function will serve as a basis for the discussion of the SCA–methods in the sequel.

2.2.1 SCA–W

In their search for a generalization of PCA, Millsap and Meredith (1988) started from the position that PCA is essentially a method that generates linear combinations (components) of the variables that explain as much variance as possible in these variables. Furthermore, they paired this position with the concept of stationary (invariant) component weights. If the same component weights are used in each group, and the variables are the same for each group, then the components have the same meaning in each group, provided that the weights matrix is used for interpretation of the components. The resulting method, called SCA–W, (Simultaneous Components Analysis with an invariant Weights matrix), generalizes PCA by computing one matrix of weights, defining components with the same meaning across groups, such that the total amount of variance explained by these components is a maximum. Millsap and Meredith (1988) suggested a gradient method to find the optimal weights in SCA–W, but this method appeared to have certain disadvantages, which can be avoided by adopting an alternating least squares algorithm (ALS) for SCA–W, as proposed by Kiers and Ten Berge (1989) and implemented by Kiers (1990).

Starting from function (2.1), the step towards SCA–W is very small. For each population the same weights matrix $\mathbf{W}$ is used, so SCA–W minimizes

$$f_w(\mathbf{W},\mathbf{P}_1,...,\mathbf{P}_p) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{W}\mathbf{P}_k' \|^2. \quad (2.2)$$

It should be noted that the single weights matrix will, in general, not lead to standardized component scores in all groups. In case standardization is desired, one should employ the relaxed constraint that

the weights matrices must be columnwise proportional. This, however, leads to the same optimization problem, and is therefore not discussed any further.

2.2.2 SCA-P

In search of a generalization of PCA, Levin (1966) capitalized on the property that PCA produces a loadings matrix **A** such that **AA**′ resembles the correlation matrix of the variables. Accordingly, he suggested a generalized method based on carrying out a PCA on the average of the correlation matrices.

Levin's idea for a generalization of PCA to two or more groups will now be related to the idea of keeping the pattern matrix **P** fixed across groups. In the definition of SCA-P, the solutions for the different groups are constrained in such a way that the pattern matrices are equal, or equivalently, perfectly congruent columnwise across all groups. Thus, SCA-P minimizes

$$f_p(\mathbf{W}_1,\ldots,\mathbf{W}_p,\mathbf{P}) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{W}_k\mathbf{P}' \|^2, \tag{2.3}$$

over $\mathbf{W}_1,\ldots,\mathbf{W}_p$ and **P**. Function (2.3) reaches a minimum when the matrix **P** ($m \times q$) contains the first q eigenvectors of the correlation matrix of $(\mathbf{Z}_1|\cdots\cdots|\mathbf{Z}_p)'$, see Kiers and Ten Berge (1994a). They also showed that, in case all groups have the same size n, this solution is the same as the one found with the method proposed by Levin (1966). In fact, Levin (1966) minimizes the function

$$g(\mathbf{P}) = \| \bar{\mathbf{R}} - \mathbf{P}\mathbf{P}' \|^2, \tag{2.4}$$

in which the unweighted mean of the correlation matrices is used, while SCA-P uses the correlation matrix of $(\mathbf{Z}_1|\cdots\cdots|\mathbf{Z}_p)'$, which is the weighted mean of the correlation matrices. Finally, it is worthwhile noting that at the minimum of function (2.3), $\mathbf{W}_1=\mathbf{W}_2=\ldots=\mathbf{W}_p=\mathbf{W}=\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}$, so instead of (2.3) we minimize

$$f_p(\mathbf{W},\mathbf{P}) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{W}\mathbf{P}' \|^2, \tag{2.5}$$

so SCA-P could also be named SCA-WP, because both the weights and the pattern matrix are kept equal over groups (Kiers & Ten Berge, 1994a). However, we will refer to this method as SCA-P.

2.2.3 SCA-S

In SCA-S, the explained variance is maximized over solutions in which the structure matrices are perfectly congruent columnwise across all groups. The method SCA-S can thus be viewed as a method that requires components to have, across groups, the same direction in the space of the variables.

The structure matrix for group k is defined as $\mathbf{S}_k = \mathbf{R}_k\mathbf{W}_k$. If the columns of the matrix $\mathbf{S}_k$ are required to be congruent columnwise across all groups, the matrix $\mathbf{S}_k$ should equal $\mathbf{SD}_k$ for a certain matrix $\mathbf{S}$ and for a certain diagonal matrix $\mathbf{D}_k$, $k=1,...,p$. Therefore, SCA-S can be described mathematically as the method that minimizes function (2.1), subject to the constraint that

$$\mathbf{W}_k = \mathbf{R}_k^{-1}\mathbf{SD}_k, \tag{2.6}$$

for certain matrices $\mathbf{S}$ and (diagonal) $\mathbf{D}_k$, assuming that $\mathbf{R}_k$ is nonsingular, $k=1,...,p$. Rewriting function (2.1), using constraint (2.6), gives

$$f_s(\mathbf{S},\mathbf{D}_1,...,\mathbf{D}_p,\mathbf{P}_1,...,\mathbf{P}_p) = \sum_{k=1}^{p} \| \mathbf{Z}_k - \mathbf{Z}_k\mathbf{R}_k^{-1}\mathbf{SD}_k\mathbf{P}_k' \|^2. \tag{2.7}$$

Kiers and Ten Berge (1994a) have presented an ALS algorithm that minimizes function (2.7).

2.3 A SIMULTANEOUS FACTOR ANALYSIS METHOD

2.3.1 Simultaneous Factor Analysis in Several Populations, using a Maximum Likelihood estimation procedure (SIFASP-ML)

Jöreskog (1971) proposed a factor analysis model for the

simultaneous analysis of two or more groups. Let $\mathbf{x}_k = (x_{k1},...,x_{km})'$ be a vector of order $m\times1$, representing the variable scores in group k, $k=1,...,p$. Jöreskog regards $\mathbf{x}_k$ as a random vector with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$ $(m\times m)$. It is assumed that a factor analysis model holds in each population, so that $\mathbf{x}_k$ can be accounted for by q common factors $\boldsymbol{\xi}_k = (\xi_{k1},...,\xi_{kq})'$ and m unique factors $\boldsymbol{\delta}_k = (\delta_{k1},...,\delta_{km})'$, as

$$\mathbf{x}_k = \boldsymbol{\mu}_k + \boldsymbol{\Lambda}_k\boldsymbol{\xi}_k + \boldsymbol{\delta}_k, \tag{2.8}$$

with $\mathcal{E}(\boldsymbol{\xi}_k)=0$, $\mathcal{E}(\boldsymbol{\delta}_k)=0$ and where $\boldsymbol{\Lambda}_k$ is a matrix of factor loadings $(m\times q)$. The usual factor analysis assumptions then imply that

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k\boldsymbol{\Phi}_k\boldsymbol{\Lambda}_k' + \boldsymbol{\Theta}_k, \tag{2.9}$$

where $\boldsymbol{\Phi}_k$ is the $q\times q$ factor covariance matrix (covariance matrix of $\boldsymbol{\xi}_k$) and $\boldsymbol{\Theta}_k$ is the $m\times m$ diagonal covariance matrix of unique factors $\boldsymbol{\delta}_k$ in group k, $k=1,...,p$. In the sequel, only deviation scores will be considered. Therefore, the vector $\boldsymbol{\mu}_k$ is specified to be the null vector.

In simultaneous factor analysis covariance matrices of two or more groups are analyzed. This means that, for each set of two or more groups, three matrices have to be estimated: $\boldsymbol{\Lambda}_k$, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}_k$, $k=1,...,p$. The simultaneous factor analysis model may specify that certain parameters in the matrices $\boldsymbol{\Lambda}_k$, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}_k$ have assigned values (fixed parameters) and that some set of unknown elements in the matrices $\boldsymbol{\Lambda}_k$, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}_k$ are the same for all k (equality constraints). The most common situation, which is the one investigated in the present study, is when the whole factor pattern $\boldsymbol{\Lambda}_k$ is assumed to be invariant across groups. So, for the present study, the model described in (2.9) is modified to

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}\boldsymbol{\Phi}_k\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_k. \tag{2.10}$$

Let $\mathbf{C}_k$ be the sample covariance matrix of group k with n_k–1 degrees of freedom. If we assume that $\mathbf{x}_k$ has a multinormal distribution it follows that the matrix $\mathbf{C}_k$ has a Wishart distribution with covariance matrix $\boldsymbol{\Sigma}_k$ and n_k degrees of freedom. The maximum likelihood estimates for the matrices $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Theta}_k$, $k=1,...,p$ are obtained by minimizing the function

$$F = \sum_{k=1}^{p} \left(\frac{n_k}{N} \right) F_{ML}(\mathbf{C}_k, \mathbf{\Sigma}_k), \tag{2.11}$$

where F_{ML} is the log likelihood function

$$F_{ML}(\mathbf{C}_k, \mathbf{\Sigma}_k(\theta_j)) = \log|\mathbf{\Sigma}_k(\theta_j)| + \mathrm{tr}(\mathbf{C}_k \mathbf{\Sigma}_k(\theta_j)^{-1}) - \log|\mathbf{C}_k| - m, \tag{2.12}$$

where the matrix $\mathbf{\Sigma}_k(\theta_j)$ is the estimate of the sample covariance matrix $\mathbf{C}_k$ using model parameters θ_j. In order to fit the model to the data in the ML–sense it is usually necessary that the model is identified. How to make sure that the model is identified will be discussed in the next subsection.

2.3.2 Identification of parameters

Suppose that the matrix $\mathbf{\Lambda}$ in the model, presented in equation (2.10), is replaced by a matrix $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{T}^{-1}$ and $\mathbf{\Phi}_k$ is replaced by $\mathbf{\Phi}_k^* = \mathbf{T}\mathbf{\Phi}_k\mathbf{T}'$, $k=1,...,p$, where $\mathbf{T}$ is an arbitrary nonsingular matrix of order $q \times q$. Then the estimated matrix $\hat{\mathbf{\Sigma}}_k$ remains the same so that function (2.11) is unaltered. Because the matrix $\mathbf{T}$ has q^2 independent elements, at least q^2 independent conditions must be imposed on the parameters in the matrices $\mathbf{\Lambda}$ and $\mathbf{\Phi}_k$, $k=1,...,p$, to make these uniquely identified. To ensure identification (in the present situation), in each column of the loadings matrix $\mathbf{\Lambda}$ (of order $m \times q$), $q-1$ elements are set to zero and one element is given the value of one. The ones in the columns must be put on different rows. Specifically, this means that a $q \times q$ block of the matrix $\mathbf{\Lambda}$ is defined to be the identity matrix (ID–block). The rows of the ID–block may be placed arbitrarily among the rows of the matrix $\mathbf{\Lambda}$. It must be noted, however, that some choices of the position of the rows of the ID–block in the matrix $\mathbf{\Lambda}$ may lead to suboptimal solutions of SIFASP–ML. In the present experiments, the rows of the ID–block are spread across the matrix $\mathbf{\Lambda}$. Further attention will be given to this topic in the description of the first experiment in which SIFASP–ML is used (Chapter 6), when discussing specific identification constraints. The remaining elements of the matrix $\mathbf{\Lambda}$ are completely free. The matrix $\mathbf{\Phi}_k$ is restricted to be symmetrical and the matrix $\mathbf{\Theta}_k$ is set to be diagonal in each group.

In SIFASP–ML, the identification constraints are merely employed to ensure that the ML fitting procedure works. Once a solution is obtained, we leave it free to rotate to any more favorable position.

In the present study, Jöreskog and Sörbom's structural equation modeling program LISREL 8 (1993) will be used to estimate the parameters in the SIFASP model, using the maximum likelihood estimation method (SIFASP–ML).

2.4 ROTATIONS USED TO FACILITATE INTERPRETATION OF THE SOLUTIONS, FOUND WITH THE METHODS OF ANALYSIS

The methods SCA–P, SCA–W, SCA–S and SIFASP–ML are, just as PCA, insensitive to both orthogonal and oblique rotations (nonsingular transformations) of the pattern, weights (not defined for SIFASP–ML) and structure matrices. For the four different methods, three different rotational techniques are used.

The solutions, found by the methods SCA–P and SIFASP–ML, are rotated using the Varimax rotation (Section 2.4.1). Note that the solutions, found by the methods SCA–P and SIFASP–ML, both contain one pattern matrix and p structure matrices. The Varimax rotated pattern matrix contains as many near zero and large elements as possible, and as few as possible elements in between. Applying this rotation to the pattern matrix, found by the methods SCA–P and SIFASP–ML, therefore generally increases the interpretability of the resulting factors, compared to the unrotated factors. The columns of the pattern matrix are rescaled to unit length before and after rotation, which makes the complete procedure the Harris and Kaiser's (1964) Independent Cluster rotation (HKIC, see Kiers & Ten Berge, 1994b).

To consider the rotation, used for SCA–W, it must firstly be noted that comparison of the solutions, found by the different methods, with the true model will primarily be based on the pattern and the structure matrices found, and *not* on the weights matrix. The reasons for not using the weights matrix in the comparison will be explained in Chapter 3. For

the solution, found by SCA–W, one may think a straightforward procedure to apply would also be the HKIC rotation, because this procedure can be seen as a method to rotate the pattern and the weights matrix simultaneously to a simple pattern and simple weights, respectively (Kiers & Ten Berge, 1994b). However, this only holds when the HKIC is applied to a single group. When applied to the solution found by SCA–W, that is, one weights matrix and p pattern matrices, the p pattern matrices are not rotated to a simple pattern. Because the pattern and the structure matrices of the solutions, found with the different methods of analysis, will be used for comparing the methods of analysis, and *not* the weights matrix, this rotation is not satisfactory in the present context.

In order to give SCA–W an opportunity that is comparable to the opportunity given to the other methods for simplifying the solution found, it is necessary to use a rotational procedure that rotates all the pattern matrices, found by SCA–W, to simple form, instead of the weights matrix. A rotational procedure that accomplishes this was proposed by Hakstian (1976, see Section 2.4.2).

As far as SCA–S is concerned, one may apply the Varimax rotation to the structure matrix, because this rotation, among other things, aims at finding some large structure elements, which facilitates interpretation. However, interpretation is facilitated even more by having components that are strongly related to certain non–overlapping subsets of variables. Therefore, in the present study, a rotational method that accomplishes this, devised by Kiers and Ten Berge (1994a), was chosen for the rotation of the structure matrix, found by SCA–S (Section 2.4.3).

2.4.1 HKIC Rotation of the SCA–P and the SIFASP–ML solution

For SCA–P, a Varimax rotation is performed on the normalized pattern matrix, found by minimizing (2.3), using a matrix $\mathbf{T}_V$, resulting in a rotated pattern matrix $\mathbf{P}_V = \mathbf{P}(\text{diag}(\mathbf{P}'\mathbf{P}))^{-\frac{1}{2}}\mathbf{T}_V$. To arrive at a correctly scaled pattern matrix for each group, $\mathbf{P}_V$ is multiplied by a diagonal matrix $\mathbf{D}_k = (\text{diag}(\mathbf{W}_V'\mathbf{R}_k\mathbf{W}_V))^{\frac{1}{2}}$, where $\mathbf{W}_V$ is the rotated weights matrix $\mathbf{W}_V = \mathbf{W}\mathbf{T}_V$, which in turn is rescaled for each group as $\mathbf{W}_k = \mathbf{W}_V\mathbf{D}_k^{-1}$. The rotated

structure and component correlation matrices for each group are then calculated as $\mathbf{S}_k = \mathbf{R}_k\mathbf{W}_k$ and $\mathbf{\Phi}_k = \mathbf{W}_k'\mathbf{R}_k\mathbf{W}_k$, respectively, the usual expressions for obliquely rotated components (see Section 1.2).

For all solutions found with SIFASP-ML we have

$$\hat{\mathbf{\Sigma}}_k = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Phi}}_k\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Theta}}_k. \tag{2.13}$$

A Varimax rotation is performed on the orthonormalized pattern matrix $\hat{\mathbf{\Lambda}}(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{-\frac{1}{2}}$, using a matrix $\mathbf{T}_V$. Next, the associated factor covariance matrix is scaled to unit diagonal by pre- and postmultiplication with $\mathbf{D}_k = (\text{diag}(\mathbf{T}_V'(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{\frac{1}{2}}\hat{\mathbf{\Phi}}_k(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{\frac{1}{2}}\mathbf{T}_V))^{-\frac{1}{2}}$. The resulting matrix with factor correlations is given as $\hat{\mathbf{\Phi}}_k^* = \mathbf{D}_k\mathbf{T}_V'(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{\frac{1}{2}}\hat{\mathbf{\Phi}}_k(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{\frac{1}{2}}\mathbf{T}_V\mathbf{D}_k$ and the associated rotated pattern matrix is $\hat{\mathbf{\Lambda}}_k^* = \hat{\mathbf{\Lambda}}(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Lambda}})^{-\frac{1}{2}}\mathbf{T}_V\mathbf{D}_k^{-1}$. The rotated structure matrix, containing correlations between the variables and the factors, for the SIFASP-ML solution is then calculated as $\hat{\mathbf{S}}_k^* = \hat{\mathbf{\Lambda}}_k^*\hat{\mathbf{\Phi}}_k^*$.

2.4.2 Hakstian's Rotation of the SCA-W solution

Hakstian's rotation procedure maximizes the sum of Varimax functions for the p pattern matrices, using the same rotation matrix for all p pattern matrices. Before the maximization, the supermatrix $\mathbf{P}_o$, containing the p pattern matrices one below the other, is made columnwise orthonormal as a form of standardization of the method. Hakstian's rotation method applied to such a supermatrix is a direct generalization of the HKIC procedure. Let $\mathbf{T}_H$ be the rotation matrix. Then $\mathbf{P}_H = \mathbf{P}_o\mathbf{T}_H$ is the supermatrix containing all rotated pattern matrices one below the other. The rotated weights matrix, structure matrix and component correlation matrix are then calculated as $\mathbf{W}_H = \mathbf{W}\mathbf{T}_H\mathbf{D}_H$, $\mathbf{S}_k = \mathbf{R}_k\mathbf{W}_H$ and $\mathbf{\Phi}_k = \mathbf{W}_H'\mathbf{R}_k\mathbf{W}_H$, respectively, with $\mathbf{D}_H = (\text{diag}(\mathbf{W}_H'\mathbf{R}_k\mathbf{W}_H))^{-\frac{1}{2}}$.

2.4.3 Simple Structure Rotation of the SCA-S solution

For SCA-S, the rotational freedom is exploited in a simple structure rotation technique, designed by Kiers and Ten Berge (1994a). The idea is to find a transformation matrix $\mathbf{T}_S$ that yields an optimal number of high correlations of the components with certain subsets of variables, across

all sets. To achieve this, the sums of certain subsets of squared elements of the rotated structure matrix are maximized, searching for non–overlapping subsets of variables, each associated with a different component. Kiers and Ten Berge (1994) have solved the maximization problem by using a monotonically convergent ALS algorithm. The rotated structure matrix is given as $\mathbf{S}_S = \mathbf{S}\mathbf{T}_S$. Using a diagonal matrix $\mathbf{D}_k = (\text{diag}(\mathbf{S}_S'\mathbf{R}_k^{-1}\mathbf{S}_S))^{-\frac{1}{2}}$, the rotated weights matrix, pattern matrix and component correlation matrix are then calculated as $\mathbf{W}_k = \mathbf{R}_k^{-1}\mathbf{S}_S\mathbf{D}_k$, $\mathbf{P}_k = \mathbf{S}_S(\mathbf{S}_S'\mathbf{R}_k^{-1}\mathbf{S}_S)^{-1}\mathbf{D}_k^{-1}$ and $\mathbf{\Phi}_k = \mathbf{W}_k'\mathbf{R}_k\mathbf{W}_k$, respectively.

2.5. MEASURES OF FIT FOR THE SCA–METHODS AND SIFASP–ML

The fit measures used in the SCA–methods and the ML–methods, used in the present study, are not comparable, because the variants of the two types of methods currently used pertain to completely different aspects of fit. The SCA–methods presented maximize the amount of explained variance of the variables. Consequently, the amount of explained variance of the variables is used as a measure of fit. The ML–methods presented, however, try to maximize the likelihood function of the parameters in a postulated model, or equivalently, minimize the chi–square value, which is directly calculated from the likelihood function (2.11).

For each SCA–method, the calculation of the variance explained by each component separately and the variance explained in each group will be presented. The amount of explained variance is defined as "The total amount of variance to be explained minus the residual variance". For SIFASP–ML, the chi–square measure is presented. Other fit measures for SIFASP–ML will be presented in Section 4.3, together with the descriptions of how these measures will be used as indicators of the optimal dimension.

2.5.1 Explained variances of the SCA–W solutions

Because the matrix $\mathbf{S}_k$ contains correlations between variables and

components, on the diagonal of $\mathbf{S}_k'\mathbf{S}_k$ are the variances explained by each component separately in group k. The explained variances per component, averaged over groups, are on the diagonal of $\sum_{k=1}^{p} n_k(\mathbf{S}_k'\mathbf{S}_k)/N$. Using (1.2), the total explained variance in group k is given by

$$\mathrm{EV}_W = \mathrm{tr}(\mathbf{W}'\mathbf{R}_k^2\mathbf{W}(\mathbf{W}'\mathbf{R}_k\mathbf{W})^{-1}) = \mathrm{tr}(\mathbf{P}_k\mathbf{\Phi}_k\mathbf{P}_k') = \mathrm{tr}(\mathbf{S}_k\mathbf{\Phi}_k^{-1}\mathbf{S}_k'). \tag{2.14}$$

The total amount of explained variance, averaged over groups is given by $\sum_{k=1}^{p} n_k \mathrm{tr}(\mathbf{S}_k\mathbf{\Phi}_k^{-1}\mathbf{S}_k')/N$.

2.5.2 Explained variances of the SCA-P solutions

Let $\mathbf{K\Lambda K}' = \sum_{k=1}^{p} n_k\mathbf{R}_k/N$ be an eigendecomposition. Then, in SCA-P we have $\mathbf{P} = \mathbf{K}_q$ and hence $\mathbf{W} = \mathbf{K}_q$. Furthermore, let $\mathbf{k}_j$ be the j^{th} column of the matrix $\mathbf{K}$, and λ_j the j^{th} eigenvalue. The variance, explained by the j^{th} component in group k is calculated as $\mathbf{k}_j'\mathbf{R}_k\mathbf{k}_j$. This is because the fitted data can be written as $\mathbf{Z}_k\mathbf{K}_q\mathbf{K}_q' = \mathbf{Z}_k\mathbf{k}_1\mathbf{k}_1' + \ldots + \mathbf{Z}_k\mathbf{k}_q\mathbf{k}_q'$, and the total explained variance for group k can be written as $n_k^{-1}\|\mathbf{Z}_k\mathbf{K}_q\mathbf{K}_q'\|^2 = n_k^{-1}\|\mathbf{Z}_k\mathbf{k}_1\mathbf{k}_1'\|^2 + \ldots + n_k^{-1}\|\mathbf{Z}_k\mathbf{k}_q\mathbf{k}_q'\|^2 = \mathbf{k}_1'\mathbf{R}_k\mathbf{k}_1 + \ldots + \mathbf{k}_q'\mathbf{R}_k\mathbf{k}_q$. The variance, explained by component j, averaged over groups is given as $(1/N)\sum_{k=1}^{p} n_k\mathbf{k}_j'\mathbf{R}_k\mathbf{k}_j = \lambda_j$. The total amount of explained variance, averaged over groups, is given by $\sum_{j=1}^{q} \lambda_j$.

2.5.3 Explained variances of the SCA-S solutions

When $\mathbf{W}_k$ is chosen such that $\mathbf{W}_k'\mathbf{R}_k\mathbf{W}_k$ has unit diagonal, and hence the matrix $\mathbf{S}_k$ contains correlations between variables and components, on the diagonal of $\mathbf{S}_k'\mathbf{S}_k = \mathbf{W}_k'\mathbf{R}_k^2\mathbf{W}_k$ are the explained variances per component for group k. The explained variance per component, averaged over groups, are on the diagonal of $\sum_{k=1}^{p} n_k\mathrm{diag}(\mathbf{W}_k'\mathbf{R}_k^2\mathbf{W}_k)/N$. The total explained variance for group k is given by (1.2), with $\mathbf{\Phi} = \mathbf{W}_k'\mathbf{R}_k\mathbf{W}_k = \mathbf{S}'\mathbf{R}_k^{-1}\mathbf{S}$, as

$$\mathrm{EV}_S = \mathrm{tr}(\mathbf{S}(\mathbf{S}'\mathbf{R}_k^{-1}\mathbf{S})^{-1}\mathbf{S}'). \tag{2.15}$$

The explained variance, averaged over groups is given by $\sum_{k=1}^{p} n_k \mathrm{tr}(\mathbf{S}(\mathbf{S}'\mathbf{R}_k^{-1}\mathbf{S})^{-1}\mathbf{S}')/N$.

2.5.4 The Chi-Square measure for SIFASP-ML

The χ^2 goodness-of-fit measure is defined as $N-p$ times the minimum of the fit function (2.11). The degrees of freedom are given as

$$df = \tfrac{1}{2}p(m)(m+1) - t, \tag{2.16}$$

where p is the number of groups, m the number of variables per group and t is the total number of independent parameters estimated in the model.

2.6 FIRST COMPARISON OF THE METHODS

2.6.1 Comparisons of PCA-sep, SCA-W, SCA-P and SCA-S

Ten Berge, Kiers and van der Stel (1992) compared SCA-W with the method, consisting of separate PCA's for each group (PCA-sep), and found that the loss in explained variance for SCA-W was relatively small. The maximal difference in percentage of explained variance between the components found with PCA-sep and SCA-W, for the one component solution, was found to be 12.5%. This difference was found in the so-called 'reversed eigenvalue case': Two covariance matrices were constructed, using 2 components, so that they had the same eigenvalues, in opposite order. For covariance matrices, constructed using three and four components (thus constructing covariance matrices of rank 3 and 4, respectively), the maximal difference in the reversed eigenvalue case, for the one dimensional solution, was found to be 10.1% and 8.58%, respectively. In a large number of randomly constructed data, the maximal discrepancies in percentages of explained variance between the methods PCA-sep and SCA-W were recorded for the one dimensional solution. For covariance matrices simulated, using 2, 3 and 4 components, these were 12.0, 8.7 and 6.8%, respectively.

Other interesting results, found by Ten Berge, Kiers and van der Stel (1992), were that SCA-W shows an average positive sampling bias in explained variance (over 14 replications), overestimating the amount of explained variance (in the population) with maximal 2.07%, and that even

at a sample size of n=50, drawn from a population of size n=2283, components were surprisingly well recovered, with average congruence values (over 14 replications) for the two components in the specific population at hand of .97 and .94, respectively.

A very important mathematical relation between the methods PCA–sep, SCA–W, SCA–P and SCA–S was discovered by Kiers and Ten Berge (1994a). They showed that the four methods PCA–sep, SCA–W, SCA–P and SCA–S are related hierarchically in that they account for (weakly) decreasing amounts of variance. Hence, one can view the methods as increasingly restrictive variants of PCA–sep. Kiers and Ten Berge formulated this as the following theorem: Let f_{sep}, f_w, f_p and f_s, for the time being, denote the total amounts of variance explained by optimal q–dimensional solutions found with the methods PCA–sep, SCA–W, SCA–P and SCA–S, respectively. Then, for any set of data $\mathbf{Z}_1,...,\mathbf{Z}_k$,

$$f_{sep} \geq f_w \geq f_p \geq f_s. \tag{2.17}$$

For the proof of this theorem, see Kiers and Ten Berge (1994a).

Ten Berge (1994) has shown that when f_w equals f_p, f_p also equals f_s. Ten Berge also showed that when we have correlation matrices $\mathbf{R}_k$ which have the same q dominant eigenvectors, the three SCA–methods and PCA–sep will explain the same amount of variance when q components are drawn. The remaining number of eigenvectors (ranging from zero to m–q) may be chosen completely different for each correlation matrix.

2.6.2 Comparison of the SCA–methods with SIFASP–ML

The method SIFASP–ML is inferior to the other methods with respect to the specifications necessary to start an analysis: Whereas the SCA–methods do not need any description of components in advance, SIFASP–ML does, in order to ensure identification (see Section 2.3.2). This is a serious drawback of the method, in the sense that one has to know part of the solution to be able to arrive at the solution. In the present study, this means that we have to give (part of) the solution away, in order to be able to test SIFASP–ML. One can, of course, reason that in analyzing scores on a set of variables (taken from two or more

groups) a researcher practically always has an expectation or theory on the basis of which to start the analysis. For this reason, SIFASP–ML is given the headstart it requires. The question whether this headstart is too much of an advantage for SIFASP–ML will be answered negatively.

CHAPTER 3

CONSTRUCTING DATA WITH KNOWN UNDERLYING FACTOR STRUCTURE

3.1 INTRODUCTION

The main purpose of this study is to compare the performance of the four methods of analysis, presented in Chapter 2, under various circumstances. All the data analyzed in this study come from simulation. That is, data are constructed by starting from some "true" model describing features of a population. From this population samples are drawn to which noise is added. In this chapter the method of data simulation is presented. In order to quantify the performance of the methods, their respective solutions will be compared with the underlying features of the simulated samples which, as said, together constitute the true model. These features are the pattern matrix, the structure matrix, the weights matrix and the factor correlation matrix. When talking about these underlying matrices, they will be given the adjective "true", because they come from the "true" model. For the different kinds of data to be simulated in this study, the true pattern, true structure, true weights and true factor correlation matrix were demanded. For the true weights matrix this appeared to be an impossibility, as will be explained in Section 3.2.2.

3.1.1 Terminology

The following definitions will be maintained throughout the present chapter:

n_k = a variable indicating sample size in group k, k=1,...,p,

$\mathbf{Y}_k$ = an $n_k \times m$ matrix of raw scores of n_k individuals on m variables,

$\mathbf{Z}_k$ = standardized version of $\mathbf{Y}_k$,

$\mathbf{R}_k$ = the associated $m \times m$ correlation matrix,

$\mathbf{F}_k = \mathbf{Z}_k\mathbf{W}_k$ = an $n_k \times q$ matrix of factor scores,

$\mathbf{\Phi}_k$ = the $q \times q$ matrix of correlations between the factors,

$\mathbf{P}_k$ = the $m \times q$ pattern matrix, containing the loadings (linear regression weights) for (re)constructing the variables from the factors,

$\mathbf{S}_k = \mathbf{P}_k\mathbf{\Phi}_k$ = the $m \times q$ structure matrix, containing the correlations between the variables and the factors,

$\mathbf{W}_k$ = an $m \times q$ component weights matrix, such that $\mathbf{F}_k = \mathbf{Z}_k\mathbf{W}_k$,

$\mathbf{E}_k$ = an $n_k \times m$ random sample matrix containing n_k realizations of a stochastic multivariate variable $\mathbf{e}_k$ (of order $m \times 1$) $\sim$ $N(0,\mathbf{D}^2_{ek})$, where

$\mathbf{D}^2_{ek}$ = an $m \times m$ diagonal matrix with in the j^{th} position on the diagonal the variance (ε^2_{jk}) of the normal distribution from which the random sample for variable j is drawn,

t = subscript indicating true population matrices (i.e. $\mathbf{P}_t$, $\mathbf{S}_t$, $\mathbf{W}_t$ and $\mathbf{\Phi}_t$).

3.2 DESCRIPTION OF THE DATA SIMULATION

For the data simulation procedure, the true pattern matrices $\mathbf{P}_{tk}$, $k=1,...,p$, and the true factor correlation matrices $\mathbf{\Phi}_{tk}$, $k=1,...,p$, have to be prespecified. Also required is a diagonal matrix $\mathbf{D}^2_{ek}$, with on the j^{th} position of the diagonal the value ε^2_{jk} $(j=1,...,m)$, which stands for the prespecified variance of the normal distribution from which the error scores for variable j are drawn. Each simulated sample k consists of a prespecified number of n_k realizations of the m–dimensional variable

$$\mathbf{y}_k = \mathbf{P}_{tk}\mathbf{f}_k + \mathbf{e}_k, \tag{3.1}$$

where $\mathbf{f}_k$ is a q–dimensional random variable with a multivariate normal distribution $N(0,\mathbf{\Phi}_{tk})$ and $\mathbf{e}_k$ is an m–dimensional random (measurement error) variable with a multivariate normal distribution $N(0,\mathbf{D}^2_{ek})$.

Each sample is generated as follows. Firstly, we construct realizations of a stochastic vector variable $\mathbf{f}_k$ with factor scores. For this the matrix $\mathbf{\Phi}^{\frac{1}{2}}_{tk}$ is multiplied by a vector $\mathbf{x}_k$, which is a stochastic multivariate variable of order $q \times 1$, governed by $N(0,\mathbf{I})$. So we have $\mathbf{f}_k = \mathbf{\Phi}^{\frac{1}{2}}_{tk}\mathbf{x}_k$ and $\mathcal{E}(\mathbf{f}_k\mathbf{f}_k') = \mathcal{E}(\mathbf{\Phi}^{\frac{1}{2}}_{tk}\mathbf{x}_k\mathbf{x}_k'\mathbf{\Phi}^{\frac{1}{2}}_{tk}) = \mathbf{\Phi}^{\frac{1}{2}}_{tk}\mathcal{E}(\mathbf{x}_k\mathbf{x}_k')\mathbf{\Phi}^{\frac{1}{2}}_{tk} = \mathbf{\Phi}_{tk}$. The n_k

realizations of the stochastic multivariate variable $\mathbf{f}_k$ are put in a matrix $\mathbf{F}_k$ with factor scores. The simulated matrix $\mathbf{F}_k$ is not of interest in itself, but is indispensable in simulating other matrices, which are of interest in the present study.

The raw scores for sample k are now calculated as

$$\mathbf{Y}_k = \mathbf{F}_k\mathbf{P}'_{tk} + \mathbf{E}_k, \tag{3.2}$$

where $\mathbf{E}_k$ contains n_k realizations of the stochastic multivariate variable $\mathbf{e}_k$. There are now two distinct ways to proceed with the data simulation. When a covariance matrix is desired, it is computed directly from the matrix $\mathbf{Y}_k$ (the scores in $\mathbf{Y}_k$ are first centered to get deviation scores). When a correlation matrix is desired, the raw scores are standardized to obtain a matrix $\mathbf{Z}_k$.

Note that the present way of data simulation is exactly the opposite from the route taken by data analysis (i.e. factor analysis). In factor analysis the variable scores are – as well as possible – compressed into factor scores, while in the present simulation process the simulated factor scores are decompressed to arrive at variable scores.

For each generated matrix with scores on the variables, the true values are demanded for the weights matrix, the pattern matrix, the structure matrix and the matrix with correlations between the factors, denoted by $\mathbf{W}_{tk}$, $\mathbf{P}_{tk}$, $\mathbf{S}_{tk}$ and $\mathbf{\Phi}_{tk}$, respectively. Of these four matrices, the matrices $\mathbf{P}_{tk}$ and $\mathbf{\Phi}_{tk}$ are chosen before the simulation. The four true matrices are required because in the present study, these true values are to be compared with the values found with the several methods of analysis. It will now be described how these matrices are chosen or derived when a covariance matrix and a correlation matrix are simulated, respectively.

3.2.1 Choosing a true pattern matrix $\mathbf{P}_{tk}$ and defining a true structure matrix $\mathbf{S}_{tk}$ when simulating a correlation or a covariance matrix

The true pattern matrix is chosen to be simple. Therefore, the pattern matrix is chosen such that each variable is constructed from only one or two factors.

When simulating a population correlation matrix, the true pattern matrix $\mathbf{P}_{tk}$ and the error coefficients in the diagonal matrix $\mathbf{D}^2_{ek}$ may not be chosen freely, because the variances of the simulated variable scores are required to be one. That is, when

$$\mathbf{z}_k = \mathbf{P}_{tk}\mathbf{f}_k + \mathbf{e}_k, \tag{3.3}$$

where $\mathbf{e}_k$ is an m–dimensional random variable with a multivariate normal distribution $N(0,\mathbf{D}^2_{ek})$, we have

$$\mathcal{E}(\mathbf{z}_k\mathbf{z}_k') = \mathcal{E}(\mathbf{P}_{tk}\mathbf{f}_k\mathbf{f}_k'\mathbf{P}_{tk}') + \mathcal{E}(\mathbf{P}_{tk}\mathbf{f}_k\mathbf{e}_k') + \mathcal{E}(\mathbf{e}_k\mathbf{f}_k'\mathbf{P}_{tk}') + \mathcal{E}(\mathbf{e}_k\mathbf{e}_k')$$

$$= \mathbf{P}_{tk}\mathcal{E}(\mathbf{f}_k\mathbf{f}_k')\mathbf{P}_{tk}' + \mathcal{E}(\mathbf{e}_k\mathbf{e}_k') = \mathbf{P}_{tk}\mathbf{\Phi}_{tk}\mathbf{P}_{tk}' + \mathbf{D}^2_{ek} = \mathbf{R}_{tk}, \tag{3.4}$$

the population correlation matrix, and hence

$$\mathbf{D}^2_{ek} = \mathbf{I}\text{–diag}(\mathbf{P}_{tk}\mathbf{\Phi}_{tk}\mathbf{P}_{tk}'). \tag{3.5}$$

In order to simulate data for which (3.5) holds, we might choose the true pattern matrices at wish (provided that the diagonal elements of $(\mathbf{P}_{tk}\mathbf{\Phi}_{tk}\mathbf{P}_{tk}')$ do not exceed 1), and take the error distribution according to (3.5).

To get insight in the signal to noise ratio for the constructed patterns, it should be noted that the variance of the 'true' part of the variables is given on the diagonal of $\mathbf{P}_{tk}\mathbf{\Phi}_{tk}\mathbf{P}_{tk}'$ and that of the error part by the diagonal elements of matrix $\mathbf{D}_{ek}$. The ratios of corresponding elements give the signal to noise ratio.

The true structure matrix is calculated as $\mathbf{S}_{tk} = \mathbf{P}_{tk}\mathbf{\Phi}_{tk}$. When the matrix $\mathbf{P}_{tk}$ pertains to nonoverlapping clusters of variables, the nonzero values of the true pattern matrix return in matrix $\mathbf{S}_{tk}$. When the off–diagonal elements of the matrix $\mathbf{\Phi}_{tk}$ are nonzero, the matrix $\mathbf{S}_{tk}$ contains nonzero values in the places where the true pattern matrix contains zero values. As a consequence, in this situation the true pattern matrix is simple and the true structure matrix is not.

The present approach to data simulation is the same as the one used in many studies comparing Principal Components Analysis with Factor Analysis (e.g. Velicer & Jackson (1990), Velicer et al. (1982), Snook & Gorsuch (1989) and Widaman (1993)). In those studies, the true correlation matrix as it holds in the population is given by

$$\mathbf{R}_{tk} = \mathbf{P}_{tk}\Phi_{tk}\mathbf{P}'_{tk} + \mathbf{U}^2_{tk}, \tag{3.6}$$

where $\mathbf{U}^2_{tk}$ is a diagonal matrix containing unique variances of the variables in the population, ensuring that the matrix $\mathbf{R}_{tk}$ has ones on the diagonal. Definition (3.6) is, in fact, equal to (3.4). The true pattern matrix always contains nonoverlapping clusters of variables (called congeneric factors by Widaman, 1993) with loadings which are less than one (mostly the values .4, .6 and .8 are chosen). The difference in data construction of the studies mentioned with the present study is that in the present study scores are constructed for all (simulated) individuals in a sample, while in the studies mentioned above, sample covariance matrices are constructed. In the present study, besides similar situations, also some more complex situations are simulated.

3.2.2 The non-uniqueness of the true weights matrix $\mathbf{W}_{tk}$

It will now be shown that, contrary to the true pattern and the true structure matrix, it is not possible to define a true weights matrix other than by arbitrary choice.

The "true scores matrix" $\mathbf{F}_k\mathbf{P}'_{tk}$ can be defined according to (3.2) with zero error. The simulated factor scores matrix $\mathbf{F}_k$ should contain linear combinations of these true scores, where the weights for these linear combinations define "true weights". Now the complete class of true weights matrices $\mathbf{W}_{tk}$, which satisfy

$$\mathbf{F}_k\mathbf{P}'_{tk}\mathbf{W}_{tk} = \mathbf{F}_k \tag{3.7}$$

will be sought. Any weights matrix satisfying (3.7) can serve as a true weights matrix. After premultiplication of (3.7) with $(\mathbf{F}'_k\mathbf{F}_k)^{-1}\mathbf{F}'_k$ (where we assume rank$(\mathbf{F}_k)$=q) we have

$$\mathbf{P}'_{tk}\mathbf{W}_{tk} = \mathbf{I}. \tag{3.8}$$

A general formula for the true weight matrices $\mathbf{W}_{tk}$ can be written as

$$\mathbf{W}_{tk} = \mathbf{P}_{tk}\mathbf{U}_k + \mathbf{P}^{\perp}_{tk}\mathbf{V}_k, \tag{3.9}$$

where the columns of $\mathbf{P}^{\perp}_{tk}$ are orthogonal to the columns of $\mathbf{P}_{tk}$, and $\mathbf{U}_k$ and $\mathbf{V}_k$ are arbitrarily chosen. It is important to note that the matrix $\mathbf{P}^{\perp}_{tk}$ is

of order $m \times (m-q)$ and the matrix $\mathbf{V}_k$ is of order $(m-q) \times q$. Substituting (3.9) into (3.8) gives

$$\mathbf{P}'_{tk}(\mathbf{P}_{tk}\mathbf{U}_k + \mathbf{P}^{\perp}_{tk}\mathbf{V}_k) = \mathbf{P}'_{tk}\mathbf{P}_{tk}\mathbf{U}_k = \mathbf{I}. \tag{3.10}$$

From this it follows that

$$\mathbf{U}_k = (\mathbf{P}'_{tk}\mathbf{P}_{tk})^{-1} \tag{3.11}$$

and the complete class of true weight matrices $\mathbf{W}_{tk}$ is now given by

$$\mathbf{W}_{tk} = \mathbf{P}_{tk}(\mathbf{P}'_{tk}\mathbf{P}_{tk})^{-1} + \mathbf{P}^{\perp}_{tk}\mathbf{V}_k. \tag{3.12}$$

With the derivation of (3.12) we have found that with a true scores matrix $\mathbf{F}_k\mathbf{P}'_{tk}$ of low rank q, any true weights matrix of full column rank q satisfying (3.12) will do. So there simply is no weights matrix which can be said to be more true than others calculated with (3.12). Because of this, no true weights matrix is present in the true model.

3.3 DEFINITION OF DATA CATEGORIES

Up till now, the simulation of scores of n_k individuals in sample k has been considered. Each sample is taken from a certain population, defined by a true pattern matrix $\mathbf{P}_{tk}$, a true structure matrix $\mathbf{S}_{tk}$ and a true correlation matrix $\mathbf{\Phi}_{tk}$. In the present study, the constructed data sets with (two or more samples) contain either samples taken from one population or samples from two or more (different) populations. These categories will be discussed in Sections 3.3.1 and 3.3.2, respectively. For both categories the percentages of explained variance, to be expected when simulated populations are analyzed with the three SCA–methods of analysis, will be discussed in Section 3.3.3. In the simulated populations the measurement error and sampling variability are excluded from the data. It will be shown that when sampling from different populations, the method SCA–W can, even when errorless data are analyzed, greatly outperform the methods SCA–P and SCA–S.

3.3.1 Sampling from one population

When samples are taken from one and the same population, the same true pattern matrix $\mathbf{P}_t$ and the same true correlation matrix $\mathbf{\Phi}_t$ underlies each sample to be simulated. So for each sample k we have $\mathbf{P}_{tk} = \mathbf{P}_t$ and $\mathbf{\Phi}_{tk} = \mathbf{\Phi}_t$. For this sort of data (from now on referred to as one–population data) all the samples have the same underlying true pattern matrix $\mathbf{P}_t$ and the same underlying true structure matrix $\mathbf{S}_t$. Furthermore, the true pattern matrix $\mathbf{P}_t$ underlying one–population data is simple (usually pertaining to nonoverlapping clusters of variables) while the true structure matrix $\mathbf{S}_t$ is usually not simple (unless the true matrix $\mathbf{\Phi}_t$ is the identity matrix).

A special case of one–population data is the situation where the true pattern matrices are not identical, but columnwise perfectly congruent across groups. In this situation, the amount of measurement error differs over the groups, but they can still be considered as coming from the same population, because it is obvious that the factors are still interpreted the same in all groups, although differing in strengths across groups.

To ensure that the columns of the two true structure matrices are perfectly congruent, it suffices to choose the matrix $\mathbf{\Phi}_{tk}$ as the identity matrix for all groups and the matrix $\mathbf{P}_{tk}$ columnwise proportional over the groups (because then we have $\mathbf{S}_t = \mathbf{P}_t$) or to choose the matrix $\mathbf{\Phi}_{tk}$ unequal to the identity matrix, but the same for all groups and choose the matrix $\mathbf{P}_{tk}$ either exactly the same for all groups (i.e. the error for each variable may differ, as long as all groups are treated equally) or choose the matrix $\mathbf{P}_{tk}$ so that within each simulated group all nonzero loadings are equal.

3.3.2 Sampling from two or more populations

When samples are taken from different populations, a different true pattern matrix $\mathbf{P}_{tk}$ is chosen for each sample to be simulated. The correlation matrix $\mathbf{\Phi}_{tk}$ may be chosen different for each sample to be

simulated, but this is not a necessity. For this reason, we will call these data two-or-more-populations data. The columns of the true pattern matrices are not proportional over groups, because then we would have one-population data. For two-or-more-populations data all the samples have different underlying true pattern matrices $\mathbf{P}_{tk}$ and different true structure matrices $\mathbf{S}_{tk}$. In the present study, the true pattern matrices $\mathbf{P}_{tk}$ underlying two-or-more-populations data are simple (nonoverlapping clusters of variables) while the true structure matrices $\mathbf{S}_{tk}$ are not simple (unless the matrix $\mathbf{\Phi}_{tk}$ is the identity matrix).

3.3.3 Explained variances in the populations

In Section 2.6.1, it was shown that the three methods of simultaneous components analysis SCA-W, SCA-P and SCA-S show a hierarchical order in the amounts of variance explained by the components found with the respective methods, which is written as

$$f_w \geq f_p \geq f_s, \tag{3.13}$$

where f stands for the variance explained and the indexes w, p and s stand for SCA-W, SCA-P and SCA-S, respectively. For the two categories of data presented, a stronger relationship between the amounts of variance explained by the three methods can be given when no error (or almost no error) has been added to the simulated scores. This will now be explained.

When one population is simulated with q factors, it is obvious that there is only one correlation matrix describing that population without error. This true correlation matrix for that population can be described as $\mathbf{R}_t = \mathbf{P}_t\mathbf{\Phi}_t\mathbf{P}_t' = \mathbf{S}_t\mathbf{\Phi}_t^{-1}\mathbf{S}_t'$. Therefore, SCA-P and SCA-S will both be able to explain exactly 100% of the variance when q components are drawn, because there is one pattern and one structure matrix perfectly describing both samples. From (3.13) it follows that SCA-W will also explain 100% of the variance. So for one-population data without error we have

$$f_w = f_p = f_s. \tag{3.14}$$

When two or more different populations are simulated the correlation

matrices describing the populations will differ to some degree. When the columns of the matrices $\mathbf{P}_{tk}$ span different spaces we have

$$f_w > f_p. \tag{3.15}$$

The proof for this strict inequality will now be given in two steps. Firstly, it will be shown that, in the error-free case, SCA-W is always able to explain 100% of the variance, whether we have one-population data or two-or-more-populations data. Secondly, it will be shown that SCA-P can not explain 100% of the variance in this case.

When population scores are simulated as described, the (errorless) data matrix $\mathbf{Z}_{tk}$ has at most rank q. Now, assume that a weights matrix $\mathbf{W}$ is used to calculate a factor scores matrix $\mathbf{F}_k = \mathbf{Z}_{tk}\mathbf{W}$, which has the same rank as the matrix $\mathbf{Z}_{tk}$. The columns of this factor scores matrix $\mathbf{F}_k$ span the same space as the columns of the scores matrix $\mathbf{Z}_{tk}$. Hence, with suitable pattern matrices $\mathbf{P}_k$, $k=1,...,p$, all the matrices $\mathbf{Z}_{tk}$ can be *perfectly* retrieved, so we have $\mathbf{Z}_{tk}\mathbf{W}\mathbf{P}_k' = \mathbf{Z}_{tk}$. This explains why always (whether we have one-population data or two-or-more-populations data) a weights matrix can be found, that is the same for all populations and has perfect fit in each population. So SCA-W will always be able to explain 100% of the variance.

It is, however, not possible to describe different true correlation matrices $\mathbf{R}_{tk}$ in two-or-more-populations data, which have at most rank q, with one pattern matrix, because the columns of the true pattern matrices $\mathbf{P}_{tk}$, making up the true correlation matrices $\mathbf{R}_{tk} = \mathbf{P}_{tk}\mathbf{\Phi}_{tk}\mathbf{P}_{tk}'$, span different subspaces. So the components found with SCA-P will always explain less than 100% of the variance.

The largest possible difference in explained variance between SCA-W and SCA-P, in the case where just two rank q correlation matrices are analyzed, is 50%. This happens when the true pattern matrices $\mathbf{P}_{t1}$ and $\mathbf{P}_{t2}$ are chosen orthogonal. Then the components found with SCA-W explain 100% of the variance and the components found with SCA-P explain 50% of the variance (see Appendix C).

For SCA-S a problem arises in this situation. When the matrices $\mathbf{R}_{tk}$, of low rank q, span different q dimensional column subspaces, it is impossible to define a q dimensional structure matrix $\mathbf{S}$ that is the same

for all $\mathbf{R}_{tk}$, because this matrix $\mathbf{S}$ has to satisfy $\mathbf{S} = \mathbf{R}_{tk}\mathbf{W}_k$. Therefore, the solution of SCA–S for these matrices $\mathbf{R}_{tk}$ is not defined and nothing can be said about the amount of variance explained by the solutions found with SCA–S. These solutions do not exist. This is not surprising because for SCA–S it was assumed that the matrix $\mathbf{R}_k$ is nonsingular, k=1,...,p.

3.4 A MEASURE FOR THE SIMILARITY OF POPULATIONS

It has been described above how samples of simulated data can be drawn from the same or from different populations. The amount of difference between populations can be manipulated by means of the choice of the true patterns. A question that arises is: "How can we quantify the difference between the populations that underlie the simulated samples?" To answer this question, a measure, called the "Congruence Measure" (CM), which quantifies the similarity of populations, was chosen.

3.4.1 The Congruence Measure

To assess to what degree the columns of a matrix $\mathbf{A}$ lead to the same interpretation as the columns of a matrix $\mathbf{B}$, Tucker's (1951) measure of congruence was adopted. The congruence between two vectors $\mathbf{a}$ and $\mathbf{b}$ (columns of the matrices $\mathbf{A}$ and $\mathbf{B}$) is defined as

$$\phi(\mathbf{a},\mathbf{b}) = (\mathbf{a}'\mathbf{a})^{-\frac{1}{2}}\mathbf{a}'\mathbf{b}(\mathbf{b}'\mathbf{b})^{-\frac{1}{2}}. \qquad (3.16)$$

When the congruence between two vectors of factor loadings on the same variables is higher than .85, the two factors will generally be judged to be more equal than unequal (see, for instance, Haven & Ten Berge, 1977). As a generalization of this congruence measure for two vectors to a congruence measure for two matrices with q columns each, first the congruence is calculated of each column $\mathbf{a}_l$, l = 1,...,q, of the matrix $\mathbf{A}$, with each column $\mathbf{b}_l$ of the matrix $\mathbf{B}$. This results in a q×q matrix $\mathbf{M}_\phi$ with q^2 congruences. As a second step, the highest mean of the absolute values of q congruences is taken, taking only one congruence from each row and each column of $\mathbf{M}_\phi$. This mean is called the Congruence Measure (CM). The

CM will be calculated for the true pattern or structure matrices of populations as a measure to quantify the similarity of two populations.

CHAPTER 4

INDEPENDENT VARIABLES, DIMENSION INDICATORS AND SUCCESS CRITERIA

4.1 MANIPULATIONS

In the present study, many different values were chosen for the independent variables, and these values were combined in several ways to construct data that captured certain aspects of real data. In order to retain some sight on what is going on, all the values, chosen for the independent variables in the present study (Experiments 1 to 4), and the true pattern matrices, resulting from these choices, will now be summarized. With these choices, the limits of generalizability of the outcomes of the present experiments are defined. In the experiments to follow (Chapters 5 to 8), for each of which different selections of the now presented values are made, the present section will serve as a point of reference. If necessary, the reader may turn to Section 3.1 for the definitions of matrices.

4.1.1 Independent Variables

- The number of variables (m): 12 and 24; the value of the variable m refers to the number of rows in the true pattern matrix $\mathbf{P}_{tk}$ and the number of columns in the matrix $\mathbf{F}_k$;
- The number of factors (q) explaining the variables in the population: 2 and 4; the value of the variable q refers to the number of columns in the true pattern matrix $\mathbf{P}_{tk}$ and the order of the true correlation matrix $\mathbf{\Phi}_{tk}$;
- The correlations between factors ($\mathbf{\Phi}_{tk}$): 0, 0.2 or 0.4. Within each correlation matrix $\mathbf{\Phi}_{tk}$ all the correlations are the same (i.e. 0, 0.2 or 0.4);
- The sample size (n_k): 50, 100 and 150 (and 300, 500 and 1000 in

Experiment 4b); The sample size refers to the number of rows in the matrix $\mathbf{F}_k$;

- The number of samples (groups) taken (p): 2, 3 or 4; The value of p indicates how many samples are constructed with a true pattern matrix $\mathbf{P}_{tk}$, k=1,...,p, and a true correlation matrix $\mathbf{\Phi}_{tk}$, k=1,...,p;

- The choice of the pattern matrix. In all studies presented here, a limited number of different values was used for the nonzero loadings in the true pattern matrices. The elements of the pattern matrices determine the strength of the factors. For simple patterns with nonoverlapping factors, the squared loading (λ) of each variable is equal to 1 minus the error variance, so the ratio of noise to signal variances for each variable is equal to $(1-\lambda^2)/\lambda^2$, and, in terms of standard deviations, the noise–to–signal ratio (nsr) is defined as

$$nsr \equiv ((1-\lambda^2)/\lambda^2)^{\frac{1}{2}}. \tag{4.1}$$

In a pilot experiment (see Section 4.5) it was found that the most appropriate values for the nsr were .5, 1.5 and 2.5, corresponding with true loadings λ of approximately .89 , .55 and .37, respectively. In the case of overlap, the same nsr's were used for comparability. For the variables, with nonzero loadings on more than one factor, the nsr's do not correspond with the above mentioned loadings.

- Degree of overlap between the factors. Overlap between factors is defined as the situation where a variable has a true loading unequal to zero on two or more factors. In the no overlap situation each variable has only one nonzero true loading. Overlapping factors have been used in Experiment 4c only, where some variables load on two factors in the true pattern matrix instead of one;

- Method of analysis: The four methods used to analyze the constructed data sets were SCA–W, SCA–P, SCA–S and SIFASP–ML, as presented in Chapter 2.

For each condition, described by choices of n, q, $\mathbf{P}_{tk}$ and $\mathbf{\Phi}_{tk}$, ten data sets were analyzed by each method. The maximizations in the methods SCA–W and SCA–S (which require an iterative procedure) were performed six times for each of the ten replications, once starting iterations with a rationally derived matrix for the matrices $\mathbf{W}$ and $\mathbf{S}$, respectively, and

five times starting with a random matrix. Both rational starts are based on the eigendecomposition $\sum_{k=1}^{p} \mathbf{R}_k = \mathbf{K}\Lambda\mathbf{K}'$. The start *weights* matrix is then derived as $\mathbf{W} = \mathbf{K}_q\Lambda_q^{-\frac{1}{2}}$, where $\mathbf{K}_q$ is the $m \times q$ matrix containing the first q columns of matrix $\mathbf{K}$ and Λ_q is the $q \times q$ matrix containing the upper left quadrant of matrix Λ; for the rational *structure* matrix, the first q columns of $\mathbf{K}$ are taken. As random start matrices for both the weights and the structure matrices, an $m \times q$ matrix of random values between plus and minus .5 is taken, which is made columnwise orthonormal by applying Gram–Schmidt orthonormalization. The best of the six solutions was taken. The iterative procedure for the special rotation procedure of the SCA–S solution was performed 21 times for each replication; once starting iterations with a rationally derived rotation matrix $\mathbf{T}$ and 20 times starting with a random rotation matrix $\mathbf{T}$. As the rational start matrix, the varimax rotation matrix, derived when rotating the rowwise normalized structure matrix, was used; as a random start matrix for the rotation, a matrix containing values between plus and minus .5 was derived which was then normalized columnwise.

The convergence criterion for the iterative procedures in the methods SCA–W and SCA–S was determined as follows: When the difference between subsequent function values (for iteration i and i+1) is smaller than .00001 times the function value at iteration i, the program stopped iterating. A justification for this convergence criterion comes from the pilot experiment, presented in Section 4.6.

4.1.2 The true matrices underlying the constructed data

The true pattern matrices were chosen in such a way that they pertained to 2 or 4 factors with non–overlapping clusters of variables (except in Experiment 4c). The true pattern matrices, for 2 and 4 factors, and 12 variables, were rowwise rescaled versions of the matrices

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

respectively. So in the 2 factors situation there were 6 variables per factor and in the 4 factors situation there were 3 variables per factor in the true pattern matrix.

The different true pattern matrices, of which selections were used in each of the different experiments, are given in Appendix A. It can be seen that different combinations of strong and weak factors are used in choosing the true pattern matrices. Only the true pattern matrices in the 12 variables condition are given, for the 2 and 4 factor condition, respectively. In the 24 variables condition, the respective true pattern matrices from the condition of 12 variables were duplicated and stacked one below the other. For instance, the true pattern matrix for 24 variables, analogous to the true pattern matrix $\mathbf{P}_a$ given below for the 12 variables condition, is given as $\tilde{\mathbf{P}}_a = (\mathbf{P}_a' | \mathbf{P}_a')'$. In the presentation of the loadings in Appendix A, only the first two decimals are given, but the loadings used were not rounded.

While the loadings in the true pattern matrices almost always can be grouped into nonoverlapping clusters of variables (the only exception is Experiment 4c), the correlations in the true structure matrix can only be grouped into nonoverlapping clusters of variables when the matrix with true correlations between the factors is chosen as the identity matrix, because the true structure matrix is derived as $\mathbf{S}_t = \mathbf{P}_t \mathbf{\Phi}_t$. In all other cases the correlations in the true structure matrix can only be grouped into overlapping clusters of variables.

Six different true correlation matrices $\mathbf{\Phi}_t$, with correlations between the factors, were used. These were

orthogonal: $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ or $\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$,

mildly oblique: $\mathbf{\Phi}_{.2} = \begin{pmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{pmatrix}$ or $\mathbf{\Phi}_{.2} = \begin{pmatrix} 1.0 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1.0 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1.0 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1.0 \end{pmatrix}$ and

strongly oblique: $\mathbf{\Phi}_{.4} = \begin{pmatrix} 1.0 & 0.4 \\ 0.4 & 1.0 \end{pmatrix}$ or $\mathbf{\Phi}_{.4} = \begin{pmatrix} 1.0 & 0.4 & 0.4 & 0.4 \\ 0.4 & 1.0 & 0.4 & 0.4 \\ 0.4 & 0.4 & 1.0 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1.0 \end{pmatrix}$,

respectively.

4.2 DIMENSION INDICATORS FOR THE SCA-METHODS

In this section, four measures are presented that were devised in order to determine the number of components to retain, when doing a simultaneous components analysis. In the present study, these measures of fit will be called dimension indicators.

4.2.1 Additional value greater than one (KA1)

An often used rule of thumb says that a component must explain at least as much variance as one variable would, for it to be added to the solution (the Kaiser eigenvalue greater than one rule, K1). Zwick and Velicer (1986) found that in the majority of Monte Carlo studies, analyzing a single group with PCA, this rule overestimated the number of components, but also examples indicating the opposite have been presented (e.g. Cliff, 1988).

In the present study, a variant of this measure was used in determining the number of components to retain, called the "additional value greater than one"-measure (KA1). The variant is based on the

rationale that, to increase the dimensionality of the solution from q–1 to q, the incremental value of the overall percentage of explained variance (denoted as d_q) needs to be larger than the percentage of variance explained by one variable. Accordingly, the number of components indicated by the measure KA1 (denoted as q_{KA1}), is the highest value of q for which the inequality

$$d_q > 100/m \tag{4.2}$$

holds. So, for the KA1–measure, the original 'eigenvalue greater than one'–part from the K1–measure is replaced by 'additional value greater than one'. This is done, because SCA–W and SCA–S do not involve eigenvalues.

4.2.2 Parallel Analysis (PA)

Parallel analysis (Horn, 1965) is a sample–based adaptation of the population–based K1 rule. Horn (1965) noted that at the population level, the eigenvalues of a correlation matrix of uncorrelated variables would all be 1.0. When samples are generated based upon such a matrix, however, the first eigenvalues exceed 1.0 whereas the last eigenvalues are below 1.0. Horn suggested that the eigenvalues of a random sample correlation matrix of m variables should be contrasted with those of the data set in question, based upon the same sample size. Components of the matrix of interest, which corresponds to eigenvalues greater than those of the comparison random matrix, are retained. According to Zwick and Velicer (1986), "this approach integrates the reliability and data–summarizing emphasis of the population–based K1 rule and the effects of sample size." In practice, a number of comparison random matrices is used, and the eigenvalues found are averaged, to attain a more accurate assessment of the random eigenvalue. Glorfeld (1995) proposed an improvement on PA by using a considerably larger number of random correlation matrices to find the complete sampling distribution of eigenvalues for each eigenvalue extracted.

Zwick and Velicer (1986) examined the accuracy of five rules (among which were the K1 rule and PA) in determining the number of components to

retain in PCA. Of the five rules, PA gave by far the most accurate determination of the number of components. For the present study, a generalization of the PA rule to two or more samples was proposed. For a proper generalization, sets of two or more random samples have to be analyzed by each method separately. The incremental amount of explained variance for each dimension corresponds to the eigenvalue for a dimension in the one sample situation. The generalization of Glorfeld's (1995) modification was considered too laborious for the present study.

For each of the three SCA–methods, an analysis of the random comparison matrices was done. To increase the dimensionality of the solution from q–1 to q, the incremental value of overall explained variance (denoted as d_q), explained by the components found with an SCA–method of the data matrices, needs to be higher than the incremental value of overall variance explained with the components found with the same SCA–method in the analysis of the random comparison matrices. The incremental value of overall variance, from the analysis of the comparison matrices, was taken as the mean value found by each method over 50 replications (and denoted by IV_q; Incremental Value found when analyzing 50 (sets of) random comparison matrices). In formula, the number of components indicated by the measure PA (denoted as q_{PA}), is found at the highest value of q for which the inequality

$$d_q > IV_q \tag{4.3}$$

holds.

4.2.3 The Quotient of Additional values (QA)

When doing a simultaneous component analysis, choosing a q–dimensional solution, the amount of variance explained by the q components can be calculated. For the (q+1)–dimensional solution, the amount of variance explained by these q+1 components will be at least as large as the amount of variance explained by the q components.

The rationale behind the QA measure is that it is expected that, when a solution is calculated with one dimension higher than the correct dimension, there will be a drop in the additional amount of explained

variance, and it is expected that increasing the dimensionality further will not give comparatively larger drops in explained variance. Denoting the increase in the total percentage of variance explained when increasing the dimensionality of the solution from q–1 to q as d_q, the quotient of additional values (QA) is defined as the quotient of d_q and d_{q+1}:

$$QA_q = \frac{d_q}{d_{q+1}}, \tag{4.4a}$$

with the requirement that

$$d_q > 100/m; \tag{4.4b}$$

the value of d_q must be larger than the amount of variance, explained with one variable. If not, the value QA_q is not calculated. Due to (4.4b), the maximum number of components possible equals q_{KA1}, the number of components indicated by KA1.

The interpretation of this measure is straightforward. When the value QA_q is not considerably larger than one, adding the $(q+1)^{th}$ component increases the total amount of explained variance almost as much as adding the q^{th} component. Therefore, a value of QA_q close to one indicates there is no reason to choose q components. When, however, the QA_q has a value of, for instance, six, this means that adding the q^{th} component increases the total amount of explained variance six times as much as adding the $(q+1)^{th}$ component. Thus we can say that the profit from adding the $(q+1)^{th}$ component is much less than the profit from adding the q^{th} component, and a choice for no more than q components seems warranted.

In the present study, the QA was used as a comparative measure. That is, the dimension that received the highest QA value was chosen as the correct dimension, so q_{QA} is the value of q for which

$$QA_q = \max(QA_k),\ k=1,\ldots,q_{KA1}. \tag{4.5}$$

From the definitions of the measures QA and KA1, it is obvious that the dimension, indicated by the measure QA, will always be lower than or equal to the dimension, indicated by the measure KA1, so

$q_{QA} \leq q_{KA1}$.

For the methods PCA and SCA–P, the amount by which the total amount of explained variance increases, decreases with each component added, because the solutions found with these methods are nested. Because the solutions, found with the methods SCA–W and SCA–S, are not nested, for these methods, this conclusion cannot be drawn. In looking at the outcomes of the experiments in this study, we have seen for the methods SCA–W and SCA–S that it was possible that adding the q^{th} component showed a smaller increase in the total amount of explained variance than adding the $(q+1)^{th}$ component. This means that the measure QA has a minimum possible value of one for the methods PCA–sep and SCA–P, whereas for the methods SCA–W and SCA–S we only know that it can be smaller than one.

4.2.4 The Quotient of Differences in Additional values (QDA)

For the QA–measure, it was mentioned that, when a solution is calculated with one dimension higher than the correct dimension, a drop in the additional amount of explained variance is expected. For the 'Quotient of Differences in Additional values'–measure (QDA), it is argued that, after the significant drop in additional amount of explained variance, further decreases in additional explained variance will be relatively small.

Denoting the increase in the total percentage of variance explained by adding the q^{th} component as d_q, the QDA for component q was defined as the quotient of $d_q - d_{q+1}$ and $d_{q+1} - d_{q+2}$:

$$\mathrm{QDA}_q = \frac{d_q - d_{q+1}}{d_{q+1} - d_{q+2}}, \qquad (4.6a)$$

again, subject to requirement (4.4b): the value of d_q must be larger than the amount of variance, explained with one variable, for the same reason as for the measure QA. The rationale behind the QDA measure can now be thought of as follows. If a drop in additional amount of explained variance occurs when increasing the dimension of the solution with the correct dimension by one, as was argued, the value of d_{q+1} will be small

compared to d_q, so the difference $d_q - d_{q+1}$ will be considerably larger than zero. However, if further decreases in d_q are small when adding another dimension to the solution, the value of d_{q+2} will be close to the value of d_{q+1}, so the difference $d_{q+1} - d_{q+2}$ will be close to zero. Therefore, the solution with the correct dimension is expected to receive the highest value of the QDA measure.

Note from the remarks at the end of Section 4.2.3, that for the methods SCA–W and SCA–S, the value of the measure QDA can become negative. It was found to be possible that the absolute value of this negative value was higher than the highest positive value. To incorporate this in formula (4.6a) the absolute value of the differences in additional values was taken, so for the methods SCA–W and SCA–S, the adjusted formula

$$\mathrm{QDA}_q = \frac{|d_q - d_{q+1}|}{|d_{q+1} - d_{q+2}|} \tag{4.6b}$$

was used for the QDA measure.

In the present study, the dimension that received the highest QDA value was chosen as the correct dimension, so q_{QDA} is the value of q for which

$$\mathrm{QDA}_q = \max(\mathrm{QDA}_k),\ k=1,\ldots,q_{KA1}. \tag{4.7}$$

Again, it is clear that $q_{QDA} \leq q_{KA1}$.

4.2.5 The use of the dimension indicators in the present study

In the experiments to be reported below, interest is not so much in the values of the dimension indicators themselves, but rather in how they compare to the "true dimensionality" underlying the data. Specifically, it was inspected whether the correct dimension was indicated, or over– or underestimation of the correct dimension occurred.

4.3 DIMENSION INDICATORS FOR SIFASP-ML

In the present study, SIFASP–ML models were fitted using the LISREL program (specifically, LISREL 8E, Jöreskog & Sörbom, 1993). The reasons for using LISREL8E rather than LISREL8 are that it can handle larger problems and executes considerably faster than LISREL8.

In the evaluation of the fit of a model, the LISREL program presents several fit indices. In the present study, a selection of these fit indices was used to evaluate the fit of competing models, and thereby to determine the number of factors to retain. Therefore, the fit indices will be called dimension indicators. Besides the sequential chi–square difference test, based on the chi–square value, selected because this is the (mis)fit index mostly used, also a fit index using the error of approximation and a fit index based on cross–validation were selected, because of their promising features.

Before introducing these fit indices, let it be noted that it is not possible to use the various indices of fit, used in factor analysis, in such a way, that all researchers with experience in the field of factor analysis will agree on the number of factors to retain. Let it also be noted that many researchers deem it unwise to base the choice for the number of factors on only one of the fit measures to be described. In the present study, the values of the fit measures were used both separately (Section 4.3.1 to 4.3.4) and in conjunction with one another (Section 4.3.6) to decide on the number of factors to retain. The purpose of utilizing the fit indices in these two different ways was to assess what usage of these fit indices is preferable. The fit measures were evaluated taking into account only the solutions for which no irregularities occurred (Section 4.3.7).

For the introduction of the fit indices, chosen for the present study, some definitions are useful, additional to the definitions given in Section 3.1.1. For reasons of consistency, some of the following notations differ from what is usual in multivariate statistics (e.g., the letter **C** is used to denote the sample covariance matrix, instead of the letter **S** that is normally used, because the letter **S** is already in use to

denote the structure matrix).

$\boldsymbol{\Sigma}_o$ = the covariance matrix of the population from which a sample is drawn. It is assumed that the sample covariance matrix $\mathbf{C}$ converges in probability to the population covariance matrix $\boldsymbol{\Sigma}_o$ with increasing sample size n;

$\boldsymbol{\theta}_j$ = the vector of parameters specifying model M_j;

$\tilde{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_j)$ represents the best possible fit of the postulated model M_j to the population covariance matrix $\boldsymbol{\Sigma}_o$ in terms of the discrepancy function F_{ML};

$\hat{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)$ represents the best fit of the postulated model M_j to the sample covariance matrix $\mathbf{C}$ in terms of the discrepancy function F_{ML};

$\hat{F}_j = F_{ML}(\mathbf{C};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j))$ is the minimum value of the discrepancy function F_{ML}, based on the best fit $\hat{\boldsymbol{\Sigma}}_j$ of the postulated model M_j to the sample covariance matrix $\mathbf{C}$;

k = subscript indicating the k^{th} group;

t_j = the number of parameters estimated in model M_j;

d_j = the degrees of freedom of model M_j;

m = the number of observed variables;

p = the number of groups;

$N = \sum_{k=1}^{p} n_k$.

In the literature, for the multiple group situation ($p>1$), the fit measures encountered were all based on χ^2. In fact, three different definitions of the chi-square measure were given, by Bentler (1989), Bollen (1989), and Jöreskog and Sörbom (1993), respectively. The only difference between the three definitions lies in the weights used for the minimized function values for each group, that is, the three authors used the weights $(n_k-1)/(N-p)$, $(n_k-1)/N$, and n_k/N, respectively. Although these differences are almost negligible, because the calculation of the chi-square value is based on asymptotic theory, it was checked which of the three weights was used in the LISREL program. It turned out to be Bentler's definition (see Section 4.3.1).

For the other fit indices used in the present study, both the

definitions of the fit indices for the single group situation and generalizations to the multiple group situation will be discussed. For the fit indices RMSEA and ECVI (Sections 4.3.3 and 4.3.4), these generalizations for the multiple group situation were derived from the values presented for these fit measures by the LISREL program.

4.3.1 The Chi-square statistic

To assess the fit of a model, many researchers will first look at the chi-square value. Let $\boldsymbol{\theta}_o$ contain the values of the parameters $\boldsymbol{\theta}$ that minimize the discrepancy function $F_o = \sum_{k=1}^{p}\left((n_k-1)/(N-p)\right) F_{ML}(\boldsymbol{\Sigma}_{ko};\boldsymbol{\Sigma}_k(\boldsymbol{\theta}))$, describing the difference between $\boldsymbol{\Sigma}_{ko}$ and $\boldsymbol{\Sigma}_k(\boldsymbol{\theta})$, $k=1,...,p$. Then the model is true when $\boldsymbol{\Sigma}_{ko} = \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_o)$ for $k=1,...,p$, in other words when $F_{ML}(\boldsymbol{\Sigma}_{ko};\boldsymbol{\Sigma}_k(\boldsymbol{\theta}_o)) = 0$. However, in practice only a sample covariance matrix $\mathbf{C}_k$ based on a sample of size n_k is available. In that case, $\hat{\boldsymbol{\theta}}$ contains the values of the parameters $\boldsymbol{\theta}$ minimizing $\hat{F}_j = \sum_{k=1}^{p}\left((n_k-1)/(N-p)\right) F_{ML}(\mathbf{C}_k;\boldsymbol{\Sigma}_k(\boldsymbol{\theta}))$, $k=1,...,p$. To test whether a model M_j is valid, under the assumption that the model is true in the population, the chi-square statistic

$$\chi_j^2 = (N-p)\hat{F}_j = \sum_{k=1}^{p}(n_k-1)\, F_{ML}(\mathbf{C}_k,\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)) \qquad (4.8)$$

can be used (cf. Bentler, 1989, p. 222–223), where F_{ML} is the discrepancy function for ML-estimation

$$F_{ML}(\mathbf{C}_k,\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)) = \log\,|\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)| + \mathrm{tr}(\mathbf{C}_k\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)^{-1}) - \log\,|\mathbf{C}_k| - m, \qquad (4.9)$$

and $\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)$ is the estimate of the population covariance matrix using model parameters $\hat{\boldsymbol{\theta}}_j$. The degrees of freedom are given as

$$df = \tfrac{1}{2}p(m)(m+1) - t_j.$$

For the maximum likelihood estimation method with discrepancy function $F_{ML}(\mathbf{C}_k;\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}))$, the chi-square measure is distributed asymptotically as a central chi-square distribution under the assumption of multinormality of the observed variables, if a covariance matrix is analyzed and under the null hypothesis that the model is exactly true. If a correlation matrix

is analyzed with ML, the χ^2–distribution applies when the model is scale–invariant and diag($\hat{\mathbf{\Sigma}}$)=diag($\mathbf{R}$) (Jöreskog & Sörbom, 1993, p. 32). In the present study, the models used were not scale–invariant. Therefore, the chi–square measure was only used as a descriptive measure for the badness of fit of a model, rather than as a measure for testing whether a model was valid, as will be described next.

4.3.2 The Sequential Chi–square Difference Test (SCDT)

In the present study, the sequential chi–square difference test (SCDT) (see, for instance, Steiger, Shapiro & Browne, 1985) was used to ascertain the "best" value for the number of common factors to retain. In this procedure, two nested models are compared by treating the difference of their chi–square test statistics as a chi–square statistic with degrees of freedom equal to the difference between the degrees of freedom for the individual chi–squares. Here, the basic idea is to interpret the difference chi–square as an "improvement index". If the difference chi–square is significant (using α=.05), it indicates that a significant improvement in the fit of the model has been produced by increasing the number of free parameters, i.e. adding one (or more) factors. The SCDT is applied repeatedly to compare all the nested models (from 1 to 6 factors), adding one factor at a time.

In the present study, in many instances only a subset of the models (with one to six factors) was present, and the SCDT could therefore not always be done. This problem was handled as follows. If the model with one factor did not fit (so no chi–square value was present), the model with two factors was taken as the first model and it was inspected whether this model had acceptable fit ($P(\chi^2_{df}<x)>.05$). If the model with two factors did not fit either, the model with three factors was inspected for acceptable fit, and so on. From the first model that fitted on, the SCDT was used. When a next model that did not fit was encountered, of course, again no SCDT could be done. In such a case, the first subsequent model with a larger number of factors was taken and the difference in chi–square values between the two models was tested for

significance.

4.3.3 Root Mean Square Error of Approximation (RMSEA)

Steiger's (1990; see also Steiger & Lind, 1980) Root Mean Square Error of Approximation (RMSEA) is a measure for the error of approximation per degree of freedom for the model. For the derivation of this measure, it is assumed that the postulated model is not valid, so the null hypothesis is false. If a postulated model is not valid, so if we have $\boldsymbol{\Sigma}_o \neq \boldsymbol{\Sigma}(\boldsymbol{\theta}_o)$ or $F_o > 0$, the statistic $\hat{F}_j$ in (4.8) asymptotically has a non–central chi–square distribution with d_j degrees of freedom and non–centrality parameter λ, defined as $\lambda = (n-1)F_o$. For the single group situation, the RMSEA is defined as

$$\text{RMSEA} = (F_o/d_j)^{\frac{1}{2}}, \tag{4.10}$$

where $F_o = F(\boldsymbol{\Sigma}_o, \tilde{\boldsymbol{\Sigma}}_j)$ can be any fit function. The sample discrepancy function value $\hat{F}_j$ is an appreciably biased estimator of F_o, the discrepancy due to approximation, which is therefore not suited as an estimate of F_o. Because $(n-1)\hat{F}_j$ has, according to the assumption made above, an asymptotic noncentral chi–square distribution with noncentrality parameter $\lambda=(n-1)F_o$ and d_j degrees of freedom, the expected value of $\hat{F}_j$ may be approximated as

$$\mathcal{E}(\hat{F}_j) \approx F_o + d_j/(n-1). \tag{4.11}$$

Consequently, a less biased point estimator of the discrepancy due to approximation F_o, would be given by $\hat{F}_o = \hat{F}_j - d_j/(n-1)$ (cf. McDonald, 1989). This statistic can have negative values, therefore

$$\max\left(\hat{F}_j - d_j/(n-1),\ 0\right)$$

was used as a point estimate for the discrepancy due to approximation. A point estimate of the RMSEA, for model M_j, is given by

$$\text{RMSEA}_j = (\hat{F}_o/d_j)^{\frac{1}{2}} = \left(\max\left(\hat{F}_j/d_j - 1/(n-1),\ 0\right)\right)^{\frac{1}{2}}, \tag{4.12}$$

(Browne & Cudeck, 1992). The generalization of the point estimate of the RMSEA for model M_j to two or more groups is given by

$$\text{RMSEA}_j = (\hat{F}_o/d_j)^{\frac{1}{2}} = \left(\max\left(\hat{F}_j/d_j - 1/(N-p),\ 0\right)\right)^{\frac{1}{2}}. \qquad (4.13)$$

Browne and Cudeck (1992) concluded on the basis of their practical experience that "a value of the RMSEA_j of about 0.05 or less would indicate a *close fit* of the model in relation to the degrees of freedom. This value is based on subjective judgment. It cannot be regarded as infallible or correct, but is more reasonable than the requirement of exact fit with the RMSEA = 0.0. We are also of the opinion that a value of about 0.08 or less for the RMSEA_j would indicate a reasonable error of approximation and would not want to employ a model with a RMSEA_j greater than 0.1" (p. 239). So, according to Browne and Cudeck, there lies a grey area between the values 0.05 and 0.1.

For the present study, the model (number of factors) with the lowest RMSEA_j, which had to be lower than 0.05, was taken as the model indicated by the RMSEA_j measure. If none of the models had an RMSEA_j lower than 0.05, it was concluded that none of the models could be chosen on the basis of the RMSEA measure. If competing models had the same lowest RMSEA value (i.e. zero), the model with the lowest number of factors was seen as the one indicated by the measure RMSEA.

In order to test the quality of the solution with the correct dimension, as found by SIFASP-ML, the test for close fit, proposed by Browne and Cudeck (1992) was used. Browne and Cudeck (1992, pp. 240–241) find the point hypothesis of exact fit

H_o: RMSEA = 0.0,

of little practical use, because this null hypothesis is invariably false in practical situations. Therefore, they proposed a test for close fit, adopted in the present study, testing the null hypothesis

H_o: RMSEA $\leq$ 0.05

against the alternative

H_1: RMSEA > 0.05.

To test this hypothesis, the test statistic $(N-p)\hat{F}_j$, having a non-central chi-square distribution, is used. Let $G((N-p)\hat{F}_j|\lambda,d_j)$ be the cumulative distribution function of the noncentral chi-square distribution with

noncentrality parameter $\lambda = (N-p)F_o = (N-p)d_j(\text{RMSEA})^2$ and d_j degrees of freedom. The probability that the stochastic variable $(N-p)\underline{\hat{F}}_j$ has a value larger than the observed value $(N-p)\hat{F}_j$, is given by

$$P\left((N-p)\underline{\hat{F}}_j > (N-p)\hat{F}_j\right) = 1 - P\left((N-p)\underline{\hat{F}}_j \leq (N-p)\hat{F}_j\right)$$

$$= 1 - G\left((N-p)\hat{F}_j | (N-p)d_j(0.05)^2, d_j\right). \qquad (4.14)$$

The null hypothesis is rejected, when the p-value is less than a prespecified 'subjective' level. In the present study, for the model with the correct number of factors, the null hypothesis was tested with α=0.05.

4.3.4 Expected Cross-Validation Index (ECVI)

Cudeck and Browne (1983) suggested that the entire approach for evaluating models should be modified: "Selection methods should de-emphasize the practice of developing a model in a single sample of data, and instead should be concerned with identifying models which will perform optimally in future samples. One reasonable method for this purpose is cross-validation" (o.c., pp. 150-151). Therefore, Browne and Cudeck (1989) proposed to use an expected value of the cross-validation index (CVI) for model selection. For this index, two samples with scores on the same variables are required. One of the samples is (arbitrarily) used as a calibration sample and the other sample is used as a validation sample, denoted by $\mathbf{C}_c$ and $\mathbf{C}_v$, respectively. The CVI is defined by Cudeck and Browne (1983, p. 151) as

$$\text{CVI} = F(\mathbf{C}_v;\hat{\mathbf{\Sigma}}_c),$$

where the estimate $\hat{\mathbf{\Sigma}}_c$ is based on the calibration sample. The smaller the value of the CVI, the more a researcher can trust that his model will hold in independent, new samples, drawn from the same population.

When only a single sample is available, however, this sample has to be divided at random into two subsamples (a calibration and a validation sample), in order to calculate a CVI. To avoid this, Browne and Cudeck (1989) devised an estimator for an unconditional expected

cross–validation index (ECVI), i.e. the expected value over calibration and validation samples:

$$\text{ECVI} = \mathcal{E}_c\mathcal{E}_v(\text{CVI}) \approx \text{F}(\boldsymbol{\Sigma}_o;\tilde{\boldsymbol{\Sigma}}_j) + (m^*+t)/(n-1), \tag{4.15}$$

where $m^* = \frac{1}{2}m(m+1)$. The point–estimate for the ECVI, under the postulated model M_j, is given by

$$\text{ECVI}_j = \text{F}(\mathbf{C};\hat{\boldsymbol{\Sigma}}_j) + 2t_j/(n-1). \tag{4.16}$$

The smaller the estimate of the ECVI, the better. Note that the ECVI_j depends on the number of parameters estimated and on the sample size. With the addition of extra parameters (increase in t_j), the function value $\text{F}(\mathbf{C};\hat{\boldsymbol{\Sigma}}_j)$ decreases. The ratio $2t_j/(n-1)$ in the formula for the ECVI makes sure that adding more parameters will not automatically lead to a lower value for the ECVI_j. The ratio $2t_j/(n-1)$ is therefore called the penalty function of the ECVI_j. When looking for the number of factors to retain, the number of factors with the lowest ECVI_j can be taken (if a minimum is reached for a number of factors). Note that it is possible that the value of the ECVI_j keeps on decreasing with each factor added. Browne and Cudeck (1992, p. 247) also noted that by inspection of the 90% confidence intervals (not given here) it is often not clear what the exact number of factors should be. According to Browne and Cudeck (1992, p. 247), the purpose of the ECVI, therefore, is merely "to give a rough indication of the number of factors that can subsequently be modified by subjective judgment." Furthermore, the value of the ECVI_j for a certain model or number of factors in itself is not important, because it can only be interpreted in comparison with values for the ECVI_j, obtained for other models.

In comparing the ECVI_j with Akaike's Information Criterion (AIC, not presented), it appears that, when ML–methods are used, there exists a linear relationship between the two (a rescaled AIC equals the ECVI_j, Browne & Cudeck,, 1992, p. 243; Jöreskog & Sörbom, 1993), so both indices lead to the same order of models. Therefore, we did not use the AIC in addition to the ECVI_j in the present study. Next, the motivation for selecting the ECVI_j will be given.

Oud, Haughton and Jansen (1996) conducted a simulation study,

evaluating fifteen criteria for model selection in LISREL, among which the fit index AIC (= $ECVI_j$). They found that in the presence of overspecified, underspecified and correctly specified analysis models, the fit index AIC showed one of the highest frequencies of selection of the correctly specified model. For this reason, the fit index $ECVI_j$ was used in the present study.

For the multiple group case, the $ECVI_j$ is calculated by LISREL as

$$ECVI_j = \hat{F}_j + 2t_j/(N-p). \tag{4.17}$$

In the present study, the number of factors for which the $ECVI_j$ measure reached its minimum value was taken as the number of factors indicated by the $ECVI_j$ measure.

4.3.5 Check of fit indices for the model with the correct dimension

In addition to the use of the fit indices as dimension indicators, as presented above, it was inspected whether the model with the correct dimension satisfied the requirement of acceptable fit. For this, it was inspected whether the p–value of the chi–square measure was larger than .05, indicating reasonable fit, and it was inspected whether the p–value for the RMSEA measure was smaller than .05, which also indicates reasonable fit.

4.3.6 The use of the fit indices as dimension indicators in the present study

In the present study, the dimension indicators were used in a comparative way (compare Section 4.2.5). It was inspected whether the correct dimension is indicated, or over– or underestimation of the correct dimension occurred.

As a test of the conjunctive use of the fit indices, a combination of the fit indices was chosen, named 'COMBI'. The dimension indicated by this measure is the dimension indicated by most of the three dimension indicators, as specified in previous sections. In the case that three different dimensions are indicated by the three dimension indicators,

according to this rule, the lowest dimension is preferred.

4.3.7 Checks on the solutions of LISREL 8

Even when given an appropriate headstart (see Sections 2.3.2 and 2.6.2), the solution found with the program LISREL–ML may contain Heywood cases, show nonconvergence and/or have otherwise improper solutions (such as non–Gramian matrices Φ_k), which make the solution suspicious, because parameters should not have values that are mathematically absurd, such as negative variances.

For the solutions, found with the LISREL program, it was inspected whether any irregularities occurred. For this, the warnings and error messages in the LISREL output were used. The irregularities can be classified into two main categories:

1.) The solution did not converge. In this case no final solution was found with the LISREL program. Nonconvergence could be attributed to: a.) The program was unable to start iterations because the fitted covariance matrix was not positive definite. In this case the Two Stage Least Squares analysis (the default start) produced nonadmissible starting values; b.) The solution, when checked after the first 20 iterations, was found to be nonadmissible. The LISREL program has a built–in check on admissibility of the solution, that checks whether the matrix Λ has full column rank and no rows of only zeros, and whether the matrices Φ and Θ are positive definite; c.) The program run reached the maximum number of iterations allowed by the program; and d.) Serious numerical problems were encountered during the program run. When this happens, the program is unable to continue iterations;
2.) The converged solution was nonadmissible. The matrix theta–delta (Θ_δ) and/or the matrix Φ, arrived at when the program had converged, were not positive definite.

In the present study, all irregular solutions were left out of the analysis. For the dimension indication, however, it was inspected what the effect was of including solutions that had converged but contained

nonadmissible values, because it was suspected that in the case that either of the matrices Φ or Θ_δ was not positive definite, this could be due to a very small divergence of parameter values beyond acceptable values.

4.4 SUCCESS CRITERIA

4.4.1 The Recovery Rate (RR)

To assess to what degree the columns of a matrix **A**, found with one of the methods of analysis, lead to the same interpretation as the columns of the appropriate true matrix **B**, Tucker's congruence measure has been adopted (see the definition of the CM in Section 3.4.1). Tucker's congruence measure is also used in the definition of an important overall success criterion called the Recovery Rate (RR). The RR gives the percentage of factors retrieved by each method. For this, for each method the choice has to be made which matrix is to be used for interpreting the components or factors, that is, the pattern matrix or the structure matrix. For the methods SCA-W, SCA-P and SIFASP-ML, the pattern matrix was the matrix for which recovery of the components or factors was measured, and for the method SCA-S this was the structure matrix. The choice for these matrices comes from the fact that these are the matrices made simple in the rotation procedure for each method. However, in Experiment 1, for each method, the RR was also calculated for the other matrix (pattern or structure), and in Experiment 3, with SCA-S also the pattern matrix was used. Now the calculation of the RR will be presented.

First, the columns of the pattern matrices found (for SCA-W, SCA-P and SIFASP-ML) or the structure matrices found (for SCA-S) were permuted and reflected to find the maximum CM (as described in Section 3.4.1) for the p matrices, found by a method, and the p true matrices. In the p matrices, compared with p true matrices, the same ordering of columns was maintained. Next, from the rows of congruences, thus found for each matrix, the number of values above .85 (the threshold obtained by Haven &

Ten Berge, see Section 3.4.1) was calculated. From the number of times a component or factor is retrieved (had a congruence value with the true factor above .85), the Recovery Rate (RR) is calculated as

$$RR = 100\% \times \frac{\text{the number of retrieved factors}}{\text{the total number of factors}}.$$

4.4.2 The Difference in Factor Correlations (DFC)

In all experiments, the correlations between the true factors were specified in the data construction. In order to investigate whether the methods of analysis, used in the present study, retrieved these correlations correctly or showed a bias, all off-diagonal elements of the true matrix Φ used were subtracted from the values of the off-diagonal elements of the matrix Φ, found with each method. For each method, the matrix Φ was calculated from the weights matrix **W** (with columns permuted and/or reflected to maximize the CM as described in Section 4.4.1) and the correlation matrix **R**. The mean of the absolute values of these differences was called the Difference in Factor Correlations measure (DFC).

4.4.3 Additional criteria for the SCA-methods: Amounts of explained variance.

As explained in Section 3.3, the simulated data used in the present study can be divided into two main categories: Data coming from one population (one-population data) and data coming from two or more (different) populations (two-or-more-populations data). In the present study, we are interested in whether or not it is possible to find an indication in the results from the SCA-methods to distinguish the data from these two categories. For this, a discriminant analysis was done in Experiment 1, with as predictors the amount of variance explained by each of the SCA-methods and the method PCA-sep, and the variables sample size, number of variables and number of factors. The number of misclassifications will be presented, together with the discriminant

function.

Imagine a situation where the strengths of factors differ across samples, while the interpretation of these factors remains the same, for instance, in a longitudinal study where some factors grow weaker over time and others grow stronger. This situation illustrates that the interpretation of a factor (discussed in Section 4.4.1) and the amount of variance it explains are entirely different and unrelated concepts, which require separate treatment. Factors can be correctly recovered, while, for instance, the strengths of weak factors are gravely overestimated, compared to strong factors. To investigate this, in Experiment 3 (Chapter 7), besides looking at the RR of factors, also the *relative* strength of components found with the SCA-methods was compared with the *relative* strength of the true factors. A measure for assessing this is discussed in Section 7.3.2.

4.5 PILOT EXPERIMENT TO DETERMINE THE PREFERABLE RANGE OF NOISE TO SIGNAL RATIOS FOR CONSTRUCTED DATA, USING THE SCA-METHODS

One of the most important preliminary questions of this study was: "At what noise-to-signal ratio do the methods considered start to yield unrealistic results?" A pilot study was conducted to answer this question. This pilot study is presented in the present section.

4.5.1 Data

In all data sets there were two groups (p=2), 12 variables (=m), 2 or 4 factors (=q), and sample sizes of n=50, n=100 or n=150. The noise-to-signal ratio's (*nsr's*) were chosen as .5, 1, 1.5, 2, 2.5 and 3. These were only varied over data sets, not over the two groups within a data set nor over factors, so we had one-population data. Each *nsr* value corresponded with a true loading (rounded to two decimals) in the pattern matrices and with a reliability (i.e., squared loading) of the

constructed variables (also rounded to two decimals) as follows (see Section 4.1.1):

nsr value	0.5	1.0	1.5	2.0	2.5	3.0
true loading	.89	.71	.55	.45	.37	.32
reliability	.80	.50	.31	.20	.14	.10

In the populations we have factors with non–overlapping clusters of variables. The true correlation matrices $\mathbf{\Phi}_{tk}$ with correlations between the factors were taken as the identity matrix. So the true matrices, for 2 and 4 factors, were $\mathbf{P}_a$, $\mathbf{P}_b$ and $\mathbf{P}_c$, given in Appendix A, and $\mathbf{P}_x$, $\mathbf{P}_y$, $\mathbf{P}_z$, defined analogously to $\mathbf{P}_a$, $\mathbf{P}_b$ and $\mathbf{P}_c$, but with nonzero loadings of .71, .45 and .32, respectively. In total, there were (2×3×6=) 36 different conditions.

In each condition, the simulation of the data was replicated five times, using different starting values for the random generation of data. The presented results of the analyses with the methods under study are the mean values for these five replications. The methods of analysis used were SCA–W, SCA–P, SCA–S and (only for comparison) PCA–sep.

4.5.2 Measures

There were two measures used to answer the question asked at the beginning of this section. Firstly, it was inspected how often the correct dimension was indicated by each of the SCA–methods in the different conditions. Because this experiment was conducted in an early stage of the study, only the measure QA (which was considered the most promising measure at the time) was used. In all conditions, for each data set the solution was calculated for 2, 3 and 4 components drawn, and from these values three quotients of additional values were calculated. Using only these three values of the QA–measure had the shortcoming that, besides indicating the correct number of components, it could only overestimate the number of components when q=2 and only underestimate the number of components when q=4. However, it still served the purpose of

the present experiment, in that we were able to see at what level the results started getting unrealistic.

Secondly, it was inspected how often the components received the correct interpretation in the different conditions: "At what error level do the SCA-methods fail to retrieve the true factors?" To answer this question, the congruence measure was adopted (instead of the RR measure presented in Section 4.4.1 just a mean congruence value was used), because this experiment was conducted in an early stage of the study, at which time the congruence measure was considered the ideal measure for recovery of factors. For each condition it was inspected, for each SCA-method separately, whether the CM for the pattern and the structure matrices for the solution with the correct dimension was above or below .85.

4.5.3 Results

The first question to be answered is: "For what noise-to-signal ratio does the measure QA fail to indicate the number of true factors for the SCA-methods (and PCA)?" Here, the measure QA was used because it was, at that point in time, deemed the most reasonable and feasible dimension indicator. The results for the QA measure are given in Table 4.1, aggregated over the four methods of analysis, resulting in a maximum of 20 possible failures for each cell (5 replications for each of the 4 methods).

From Table 4.1 it can be seen that, when large noise-to-signal ratios are used (nsr=3), the number of components, indicated by the measure QA, is often overestimated for q=2, n=50 and often underestimated for q=4, for all group sizes. On the basis of this result a choice of nsr=3 for the noise-to-signal ratio (giving true pattern matrix $\mathbf{P}_z$) seems too high.

Next, the CM of the pattern and the structure matrices, found with the SCA-methods, and the true pattern and true structure matrices, respectively, for the combinations of true pattern matrices, were

Table 4.1: *The times the QA measure indicated an incorrect number of components, summed over the different methods of analysis*[*)]

		noise-to-signal ratio					
n	q	.5	1.0	1.5	2.0	2.5	3.0
50	2	0	0	0	1	6	9
100	2	0	0	0	0	0	3
150	2	0	0	0	0	0	1
50	4	0	0	10	17	17	15
100	4	0	0	0	4	17	18
150	4	0	0	0	5	14	16

*) *read table* 1 *as follows*: '0' *means the measure QA indicated the correct number of components in all* 20 *replications*; '4' *means the measure QA overestimated* (*when* $q=2$) *or underestimated* (*when* $q=4$) *the number of components* 4 *times.*

inspected. In Table 4.2 it is indicated by a "+" that the mean CM was over .85 and by a "–" that it was smaller than .85. Because the three SCA–methods all behaved approximately the same on this measure, no distinction was made between the SCA–methods. Also, there was approximately no difference in the results found for the pattern matrix and the structure matrix. Therefore, the results in Table 4.2 go for both the structure and the pattern matrix. The only exception to this was SCA–P, that attained a mean value of the CM just above .85 for the pattern matrix in the condition n=50, q=2 and nsr=3.0. Both other methods attained a CM for the pattern and the structure matrices smaller than .85, as did SCA–P with the structure matrix, explaining the minus in this condition. In Table 4.2, the conditions for which the measure CM was above and below .85 are given.

From Table 4.2, it can be seen that high noise–to–signal ratios (nsr=3), small sample sizes (n=50) and a high number of factors (q=4 as opposed to q=2), facilitate the failure of the SCA–methods of retrieving the true factors.

Table 4.2: *The conditions for which the mean value of the CM was above (+) or below (−) .85.*

		noise-to-signal ratio					
n	*q*	.5	1.0	1.5	2.0	2.5	3.0
50	2	+	+	+	+	+	−
100	2	+	+	+	+	+	+
150	2	+	+	+	+	+	+
50	4	+	+	+	−	−	−
100	4	+	+	+	+	+	−
150	4	+	+	+	+	+	−

4.5.4 Conclusion about meaningful noise levels

From the results found with the two different measures, it was concluded that, in order to span the complete range of useful and meaningful noise-to-signal ratios, the values 0.5, 1.5 and 2.5 would be used in the bulk of this study, corresponding to true loadings of .89, .55 and .37, respectively. This explains the decision (Section 4.1.1) to settle for noise-to-signal ratios of .5, 1.5 and 2.5.

4.6 PILOT EXPERIMENT TO DETERMINE THE PREFERABLE CONVERGENCE VALUE FOR THE ITERATIVE PROCEDURES IN THE METHODS SCA-W AND SCA-S

In order to assess what value to use for the convergence criterion in the iterative procedures in the methods SCA-W and SCA-S, a pilot experiment was done, using three different values for the convergence criterion. The minimum value of the convergence criterion was sought, for which the iterative procedures arrived at the global optimum most of the time (when multiple starts were used) within a reasonable amount of time.

4.6.1 Data

In all data sets there were four groups (p=4), 18 variables (=m), 3 factors (=q), and sample sizes of n=50, n=100 or n=150. Only one noise–to–signal ratio was used: nsr=1, resulting in a true loading of .71 in the true pattern matrix, so we have one–population data. The true correlation matrix $\mathbf{\Phi}_{tk}$ with correlations between the factors was taken as the identity matrix for all groups. So the true pattern matrix is given as

$$\mathbf{P}_t = .71 \times \begin{pmatrix} \mathbf{1}_6 & \mathbf{0}_6 & \mathbf{0}_6 \\ \mathbf{0}_6 & \mathbf{1}_6 & \mathbf{0}_6 \\ \mathbf{0}_6 & \mathbf{0}_6 & \mathbf{1}_6 \end{pmatrix},$$

where $\mathbf{1}_6$ is a 6×1 vector of ones and $\mathbf{0}_6$ a 6×1 vector of zeros.

The values, chosen for the convergence criterion were .0001, .00001 and .000001. In total there were (3×3=) nine conditions. In each condition the simulation of the data was replicated three times, using different starting values for the random generation of data. For each of the three replications, both SCA–W and SCA–S were run eleven times, one time starting the iterations with a rationally derived weights matrix or structure matrix for the methods SCA–W and SCA–S, respectively, and ten times starting iterations with a randomly generated matrix. In every condition the same ten random starts were used. It was assumed that the best of the eleven solutions was the global optimum. Because in this pilot experiment, it was only of interest what the effect of convergence values used was in finding the global optimum, the solutions, found with the methods SCA–W and SCA–S, were not rotated.

4.6.2 Measures

In all conditions, the two, three and four components solutions were calculated, each solution for each method consisting of eleven function values (i.e., explained variances). The highest of these eleven values was taken as the global optimum, and the other ten values were compared with this global optimum. When the difference of one of these ten values

with the global optimum was smaller than .01, the iterative procedure was considered to have arrived at the global optimum for that particular instance.

4.6.3 Results

The main question was: "For what value of the convergence criterion is the global optimum reached for most of the starts used in the iterative procedure." When summed over sample size, method and (rational and random) starts, giving a total number of starts of (2×11×3×3×3=) 594, the percentage of times the global optimum is reached by the iterative procedures (from now on called success rates) for the three convergence values .0001, .00001 and .000001, were 55%, 83% and 97%, respectively. The random starts appear to be a little more sensitive to the choice of convergence value than the rational starts, with success rates of 78%, 91% and 93% for the rational starts for the values .0001, .00001 and .000001, respectively, and success rates of 52%, 83% and 97%, for the random starts, respectively.

Sample size has almost no effect. The number of components drawn, however, does have a large effect on whether the global optimum is reached. When the correct number of components is drawn (three), the global optimum is reached in all cases, for all three values of the convergence criterion. However, the convergence criterion does affect the success of the iterative procedures when two or four components are drawn.

4.6.4 Conclusion about meaningful convergence criteria

On the basis of the results, it was concluded a convergence value of .00001 was most appropriate for the present study (see Section 4.1.1, final paragraph) . The value .0001 showed a too high a chance to not reach the global optimum, and using the value .000001, although giving higher success rates than when using the value .00001, appeared to be very time consuming, especially for the method SCA–S. It can also be seen

that using eleven starts is wasteful, given that more than 80% of the starts tend to lead to the global optimum. Hence the decision (Section 4.1.1) to use five random and one rational start seems to be adequate.

CHAPTER 5

EXPERIMENT 1, FIRST COMPARISON OF THE THREE SCA-METHODS

5.1. INTRODUCTION

In the first experiment, conducted in this study, the objective was to differentiate between the methods SCA–W, SCA–P and SCA–S. For this, the performance of the three methods on the various measures and success criteria under a variety of circumstances was investigated. For this reason, the independent variables were manipulated over wide ranges of values, although some common characteristics were maintained throughout all conditions, and values of some independent variables were kept within realistic bounds.

5.2 MANIPULATIONS OF THE DATA

The data (correlation matrices), analyzed in this experiment, were constructed as described in Section 3.2. The following values were chosen for the independent variables. In all data sets there were two samples (p=2), 12 or 24 variables (=m), 2 or 4 factors (=q) and sample sizes of n=50, n=100 or n=150. The true pattern matrices ($\mathbf{P}_a$ to $\mathbf{P}_g$) used in Experiment 1 are given in Appendix A. These are used in twelve different combinations (see Sections 5.2.1 and 5.2.2), six of which lead to one–population data and six to two–or–more–populations data. In total six different true correlation matrices $\mathbf{\Phi}_t$, with correlations between the factors were used. The factors were defined to be orthogonal ($\mathbf{\Phi}_t$=$\mathbf{I}$), mildly oblique ($\mathbf{\Phi}_t$=$\mathbf{\Phi}_{.2}$) or strongly oblique ($\mathbf{\Phi}_t$=$\mathbf{\Phi}_{.4}$; see Section 4.1.2).

Now the rationale for the choice of the specific true pattern matrices used will be given. The first three true pattern matrices $\mathbf{P}_a$, $\mathbf{P}_b$ and $\mathbf{P}_c$ define strong factors, intermediate factors and weak factors,

respectively. With these matrices the overall effect of factor strength on the dependent variables was investigated. The true pattern matrices $\mathbf{P}_d$ and $\mathbf{P}_e$ define factors of moderately different and strongly different strengths, respectively. With these matrices it was investigated whether the retrieval of weak factors was influenced by the strengths of the other factors in the population. The true pattern matrices $\mathbf{P}_f$ and $\mathbf{P}_g$, finally, define factors of equal strengths, with unequal loadings per factor, descending and ascending in strength, respectively. These matrices were used to simulate two-or-more-populations data, described in Section 3.3. The combinations of these seven different true pattern matrices used form a reasonable representation of the different situations one might expect to encounter in the "real life" research domain.

In each condition, described by choices of the number of variables in the population (NVar), the number of factors in the population (NFac), sample size (SampSize), combinations of pattern matrices (PatternComb) and combinations of correlation matrices $\mathbf{\Phi}_t$ (PhiComb), ten replications of simulated data were used. The methods of analysis used were SCA-W, SCA-P, SCA-S and (only for comparison) PCA-sep. Now the pattern combinations leading to one-population data and two-or-more-populations data will be described.

5.2.1 One-population data

In one-population data, the columns of the true pattern matrices of the two groups are perfectly congruent. The same goes for the columns of the two true structure matrices. So the CM across groups for all true pattern matrices and all true structure matrices is 1.00 (see Section 3.3.1).

The combinations of true pattern matrices $\mathbf{P}_k$ (see Appendix A) and matrices $\mathbf{\Phi}_k$, used to simulate the two groups in each data set for one-population data, were '$\mathbf{P}_a\mathbf{P}_a$', '$\mathbf{P}_b\mathbf{P}_b$', '$\mathbf{P}_c\mathbf{P}_c$', '$\mathbf{P}_d\mathbf{P}_d$', '$\mathbf{P}_e\mathbf{P}_e$', '$\mathbf{P}_a\mathbf{P}_c$' and '**I I**', '$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$' and '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$', respectively. It can be verified that all combinations indeed satisfy the definition of one-population data

mentioned above. Within each combination of true pattern matrices and correlation matrices (18 combinations in total), data sets were simulated using 12 or 24 variables, 2 or 4 factors and sample sizes of 50, 100 and 150, amounting to 216 different conditions within the category of one-population data to be analyzed.

5.2.2 Two-or-more-populations data

In two-or-more-populations data, the true pattern and the true structure matrices are *not* proportional over the groups. The scores for two simulated groups are generated differently. That is, for each group a different true pattern matrix is used.

The combinations of true pattern matrices $\mathbf{P}_k$ (see Appendix A) and matrices $\mathbf{\Phi}_k$, used to simulate data for the two groups in each data set for two-or-more-populations data are given in Table 5.1, together with the value of the CM between the two structure matrices for each combination. The CM is reported to display the similarity between the two true structure matrices in each condition. The CM for the true pattern matrices is independent of the choice of matrices $\mathbf{\Phi}_k$, so that the values of the CM are the same for all choices of matrix $\mathbf{\Phi}_k$. They can be found on the first row of Table 5.1. From Table 5.1 it can be seen that, when the correlation matrices $\mathbf{\Phi}_k$ are chosen differently for each group, the CM for the true structure matrices drops considerably below the CM for the true pattern matrices.

Within each combination of true pattern matrices and correlation matrices (30 combinations in total), data sets were simulated using 12 or 24 variables, 2 or 4 factors and sample sizes of 50, 100 and 150, amounting to 360 different conditions within the category of two-or-more-populations data to be analyzed.

5.3 DEPENDENT VARIABLES

For each data set of two groups, the amount (percentage) of variance explained was calculated for 1, 2, 3, 4, 5 and 6 components drawn. In the

Table 5.1: *The CM for the true structure matrices*[1] *for combinations of true pattern matrices* $\mathbf{P}_k$ *and matrices* $\mathbf{\Phi}_k$, *used for two-or-more-populations data*

q	PhiComb condition	$\mathbf{P}_a\mathbf{P}_g$	$\mathbf{P}_b\mathbf{P}_g$	$\mathbf{P}_c\mathbf{P}_g$	$\mathbf{P}_d\mathbf{P}_g$	$\mathbf{P}_e\mathbf{P}_g$	$\mathbf{P}_f\mathbf{P}_g$
2	**I I**	.94	.94	.94	.94	.94	.78
2	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	.94	.94	.94	.94	.92	.78
2	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	.94	.94	.94	.93	.90	.78
2	**I** $\mathbf{\Phi}_{.4}$	.87	.87	.87	.87	.87	.72
2	$\mathbf{\Phi}_{.4}$ **I**	.87	.87	.87	.85	.80	.72
4	**I I**	.94	.94	.94	.94	.94	.78
4	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	.94	.94	.94	.93	.91	.78
4	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	.94	.94	.94	.92	.88	.78
4	**I** $\mathbf{\Phi}_{.4}$	.77	.77	.77	.77	.77	.64
4	$\mathbf{\Phi}_{.4}$ **I**	.77	.77	.77	.75	.70	.64

[1]*The CM for the true pattern matrices is the same for all choices of matrix* $\mathbf{\Phi}_k$ *and is given on the first row* (**I I**) *for* $q=2$ *and* $q=4$.

4 factor condition, the explained variance for 7 components drawn was also calculated. This was done for the measures QA and QDA. In order to get the QDA value for q components, the explained variance of q+2 components (among other things) is required (see Section 4.2.4). If the maximum number of components for which the explained variance is calculated in the 4 factors condition would be 6, the maximum number of components that could be indicated by the measure QDA would be 4, so no overestimation could occur. For this reason, the amount of variance explained by 7 components was also calculated in the q=4 condition. The 7 components solution was used for the measure QA (making possible an indication of 6 components) and QDA (making possible an indication of 5 components). For the dimension indicators KA1 and PA, the 7 component solution was not of interest, because, for the success rate of the dimension indicators it was only inspected whether or not overindication occurred, regardless of how large the overindication was.

According to the description of the dimension indicators, given in Section 4.2.5, it was recorded for each data set whether under–, correct or overestimation of the number of factors occurred. For comparison, the

results of the dimension indicators for PCA-sep were also included.

For each solution with the correct dimension, the measures RR and DFC were calculated (see Section 4.4.1. and Section 4.4.2). As far as the measure RR is concerned, the main focus was on the RR for the pattern matrix for SCA-P and SCA-W, and the RR for the structure matrix for SCA-S.

The simulated data used in the present experiment could be divided into two main categories: Data coming from one population (one-population data) and data coming from two (different) populations (two-or-more-populations data). In the present experiment, we investigated whether the data from these two categories could be distinguished by using a discriminant analysis with the variables SampSize, NVar, NFac and the percentages of explained variance of the SCA-methods and PCA-sep as predictors. The number of misclassifications will be reported, together with the discriminant function.

5.4 ANALYSIS

All the results presented were analyzed with repeated measures analyses of variance. Sums of squares of main and interaction effects were calculated to give an impression of the size of the effects, and the associated averaged univariate F tests were employed for significance testing. Multivariate criteria such as Wilks' lambda were also checked and it was found that using this criterion led to similar results. So employing these other multivariate criteria would not change the conclusions presented here.

For the analyses, it had to be checked whether or not assumptions were violated. The three assumptions for the F tests in a univariate repeated measures analysis are: Independence of observations, multivariate normality and sphericity. The procedure for simulating the data, used in the present study, guarantees that the assumption of independence of observations is fulfilled. Furthermore, the analysis is fairly robust against violation of the assumption of normality (for the

present study it is obvious that this assumption is violated) and, because the α used is chosen to be very small (α=.001), little harm can be expected from such violations. Finally, to correct for possible nonsphericity, the adjustment of the number of degrees of freedom, proposed by Greenhouse and Geisser (1959), was used.

The analyses were carried out using the MANOVA module in SPSS-PC 5.01 extended. The analyses were performed on the full data set, and on the data sets for one-population data and two-or-more-populations data, separately. The between subjects variables in the analyses of the complete data set were: Population (2 conditions, one-population data and two-or-more-populations data - note that for each population condition there were 6 different conditions of PatternComb), PhiComb (3 or 5 conditions), SampSize (3 conditions; n=50, 100, 150), NVar (2 conditions; 12 and 24 variables), and NFac (2 conditions; 2 and 4 factors). In the separate analyses of the samples from one population (one-population data) and the samples from two populations (two-or-more-populations data), the between subjects variable PatternComb was included in the analysis. In all analyses, the variable Method was specified as a within subjects factor (3 conditions; SCA-W, SCA-P and SCA-S) in a repeated measures design, because each method of analysis was presented with the same data sets to be analyzed. In each of the separate analyses, only first and second order interactions among between factors were included in the analysis. Prior analyses with a full factorial model showed that there were no relevant higher order interactions present.

For the analysis of the results for the dimension indicators, PCA-sep was included as a condition of the within subjects factor, so in that case the within subjects factor had 4 conditions. In order to be able to perform a MANOVA, observations concerning correct dimension indication were aggregated over the ten replications within each condition to obtain a reasonably continuous variable: The success rate of indicating the correct number of components over ten replications. This means that for each dimension indicator for each method, for each condition a value between 0 and 1.0 was obtained. In a first analysis, the dimension indicator used was specified as a within subjects factor (4

conditions; KA1, PA, QA and QDA) in a repeated measures design. After choosing the preferred dimension indicator, only the results for this measure were used.

5.5 QUESTIONS TO BE ANSWERED BY THIS EXPERIMENT

The present experiment was conducted to get answers as to what factors play a role in the recovery of factors by the methods at hand, and which measures can best be used for determining the number of components present in a data set. These questions are specified as below. Answers will be given in separate subsections of Section 5.6, where all results will be presented. The question numbers refer to the subsections in which answers to these questions will be given. Summary conclusions from the results of this experiment will be drawn in Section 5.7.

1.) Which of the dimension indicators KA1, PA, QA and QDA can best be used by each of the SCA-methods to indicate the number of components to retain?

2a.) Which of the SCA-methods has the highest success rate on the preferable dimension indicator?

2b.) Are there interesting interaction effects between independent variables and method of analysis for the most preferable dimension indicator?

3a.) Which of the SCA-methods can be judged as best capable of recovering factors?

3b.) Are there interesting interaction effects between independent variables and method of analysis for the recovery of factors?

4a.) Which of the SCA-methods can be judged as best capable of recovering the correlations between factors?

4b.) Are there interesting interaction effects between independent variables and method of analysis for the recovery of correlations between factors?

5.) Can samples from one population be distinguished from samples from two populations?

Besides the above questions, a few questions concerning the effect of specific choices of the pattern matrices were asked.

6a.) Which SCA–method gives the highest RR, when comparing samples taken from one population, with only strong factors (condition '$\mathbf{P}_a\mathbf{P}_a$'), only intermediate factors (condition '$\mathbf{P}_b\mathbf{P}_b$'), and only weak factors (condition '$\mathbf{P}_c\mathbf{P}_c$')?

6b.) Are, for the SCA–methods, intermediate factors and weak factors blurred by strong factors, or vice versa? For this purpose, the conditions '$\mathbf{P}_d\mathbf{P}_d$' and '$\mathbf{P}_e\mathbf{P}_e$' were compared with conditions '$\mathbf{P}_b\mathbf{P}_b$' and '$\mathbf{P}_c\mathbf{P}_c$', respectively.

6c.) Is the RR for the three SCA–methods affected by the existence of small measurement error in one sample and large measurement error in another sample (condition '$\mathbf{P}_a\mathbf{P}_c$'), and which SCA–method performs best under this condition? For this purpose, results for condition '$\mathbf{P}_a\mathbf{P}_c$' were compared with those for condition '$\mathbf{P}_a\mathbf{P}_a$' and condition '$\mathbf{P}_c\mathbf{P}_c$'.

5.6 RESULTS

In the present experiment, comparisons were made with large numbers of observations in each condition, which, in many cases, caused even small differences to become significant. Because small differences have little practical value, these will be ignored. Overall, a significance level α=.001 was used. Whenever differences between dependent variables or methods are reported, or when interaction effects are said to occur, effects were significant at α=.001.

For each of the conditions of independent variables, used in the figures, the number of data sets analyzed on which the reported values of the success criteria are based, is given in Table 5.2. In the following Sections, each of the six main questions, asked in of Section 5.5, will be answered separately. At the start of each Section, the question(s) will be repeated.

Table 5.2: *The number of observations for the conditions of the independent variables on which the values in the figures are based.*

Independent Variable	all samples	n	1 population	n	2 populations	n
All		5760				
Population	1 pop	2160	1 pop	2160	2 pop	3600
	2 pop	3600				
PatternComb	See breakdown over 1 and 2 populations		$\mathbf{P}_a\mathbf{P}_a$	360	$\mathbf{P}_a\mathbf{P}_g$	600
			$\mathbf{P}_b\mathbf{P}_b$	360	$\mathbf{P}_b\mathbf{P}_g$	600
			$\mathbf{P}_c\mathbf{P}_c$	360	$\mathbf{P}_c\mathbf{P}_g$	600
			$\mathbf{P}_d\mathbf{P}_d$	360	$\mathbf{P}_d\mathbf{P}_g$	600
			$\mathbf{P}_e\mathbf{P}_e$	360	$\mathbf{P}_e\mathbf{P}_g$	600
			$\mathbf{P}_a\mathbf{P}_c$	360	$\mathbf{P}_f\mathbf{P}_g$	600
PhiComb	$\mathbf{I}\ \mathbf{I}$	1440	$\mathbf{I}\ \mathbf{I}$	720	$\mathbf{I}\ \mathbf{I}$	720
	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	1440	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	720	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	720
	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	1440	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	720	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	720
	$\mathbf{I}\ \mathbf{\Phi}_{.4}$	720			$\mathbf{I}\ \mathbf{\Phi}_{.4}$	720
	$\mathbf{\Phi}_{.4}\ \mathbf{I}$	720			$\mathbf{\Phi}_{.4}\ \mathbf{I}$	720
SampSize	$n = 50$	1920	$n = 50$	720	$n = 50$	1200
	$n = 100$	1920	$n = 100$	720	$n = 100$	1200
	$n = 150$	1920	$n = 150$	720	$n = 150$	1200
NVar	$m = 12$	2880	$m = 12$	1080	$m = 12$	1800
	$m = 24$	2880	$m = 24$	1080	$m = 24$	1800
NFac	$q = 2$	2880	$q = 2$	1080	$q = 2$	1800
	$q = 4$	2880	$q = 4$	1080	$q = 4$	1800

5.6.1 Finding the preferable dimension indicator

Question 1: "Which of the dimension indicators KA1, PA, QA and QDA can best be used by each of the SCA–methods to indicate the number of components to retain?"

For this, the four dimension indicators were compared for each of the methods separately. In Figure 5.1a to 5.1d, the percentage of times the correct number of components was indicated by the four dimension indicators is presented for each method separately. The percentages are

given, summed over all conditions taken together, and summed over each condition of each independent variable separately.

For PCA–sep, the measure QDA gave significantly higher percentages of correct dimension indications (69.4%) than the measures QA (57.4%), PA (58.4%), and KA1 (35.5%). As can be seen from Figure 5.1a, the measure QDA also has the highest percentage of correct dimension indications of the four dimension indicators, when summed over each condition of each independent variable separately.

For SCA–W, the same holds (see Figure 5.1b). The measure QDA gave significantly higher percentages of correct dimension indications (74.6%) than the measures QA (66.3%), PA (61.6%) and KA1 (45.6%), and has the highest percentage of correct dimension indications of the four dimension indicators, when summed over each condition of each independent variable separately.

For SCA–P, the measure QDA also gave significantly higher percentages of correct dimension indications (75.3%) than the measures PA (61.8%), and KA1 (51.1%). The difference with the measure QA (69.7%) is not significant. As can be seen from Figure 5.1c, the measure QDA does not always have the highest percentage of correct dimension indications of the four dimension indicators, when summed over each condition of each independent variable separately. For instance, when 12 variables are used, the measure KA1 has a higher percentage of correct dimension indications, although differences are small.

Finally, for SCA–S, the story is somewhat different. For this method, the measure QA gives the highest percentages of correct dimension indications (55.8%), but the differences with the measure QDA (51.0%) is not significant in contrast to the differences with the measures PA (43.8%) and KA1 (38.1%). The measure QDA also has significantly higher percentages of correct dimension indications than the measures PA and KA1.

In Figure 5.2, the percentage of times the correct number of components was indicated by each dimension indicator and each method, together with the percentages of times too many or too few components were indicated, are presented. The measure KA1 almost never indicated too

few components, but it did indicate too many components more often than the correct number of components for all methods except SCA–P.

The measure PA indicated the correct number of components more often than it over– or underindicated the number of components. Both over– and underindication occurred.

The measures QA and QDA indicated the correct number of components more often than they over– or underindicated the number of components. Both measures underindicated the number of components, but overindication did not occur much, except for SCA–S.

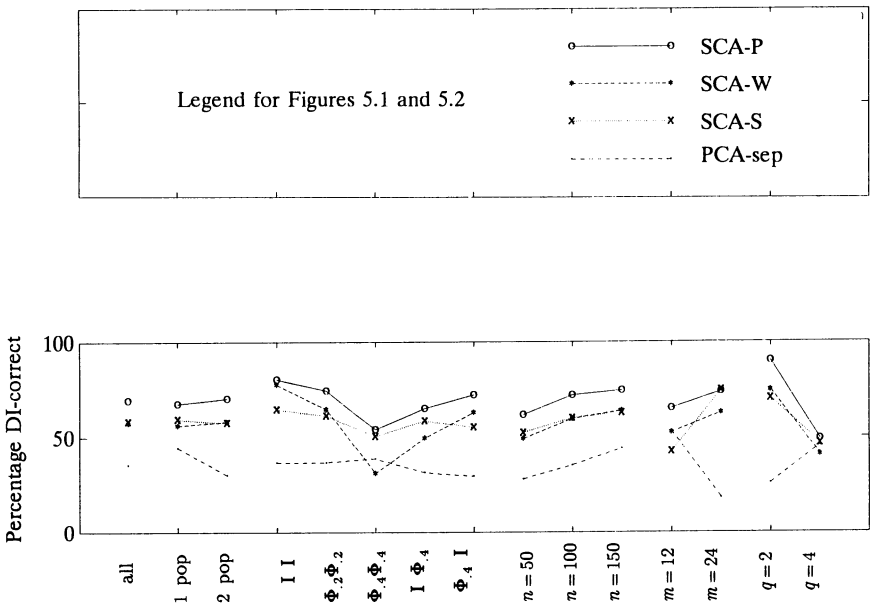


Figure 5.1a *Percentages of correct dimension indications for the method PCA–sep, averaged over each level of each independent variable, using four different dimension indicators*

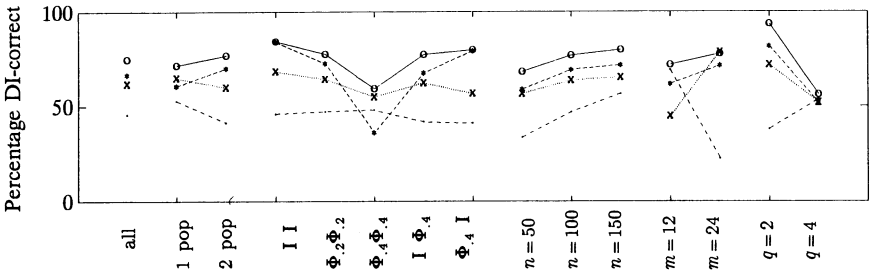


Figure 5.1b *Percentages of correct dimension indications for the method SCA–W, averaged over each level of each independent variable, using four different dimension indicators*

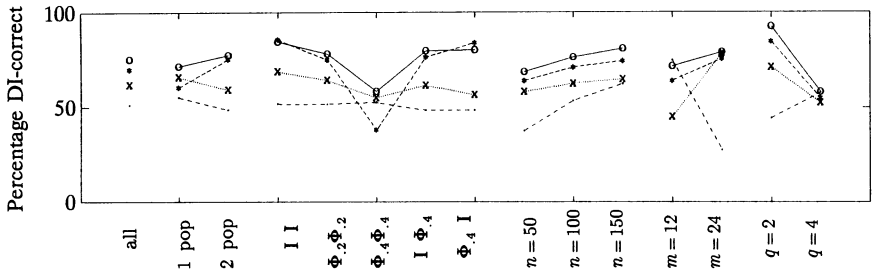


Figure 5.1c *Percentages of correct dimension indications for the method SCA–P, averaged over each level of each independent variable, using four different dimension indicators*

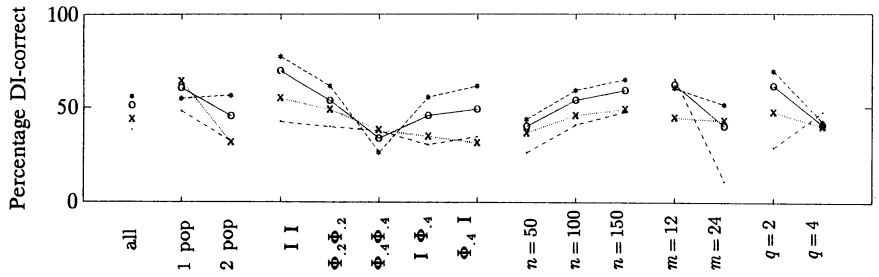


Figure 5.1d *Percentages of correct dimension indications for the method SCA–S, averaged over each level of each independent variable, using four different dimension indicators*

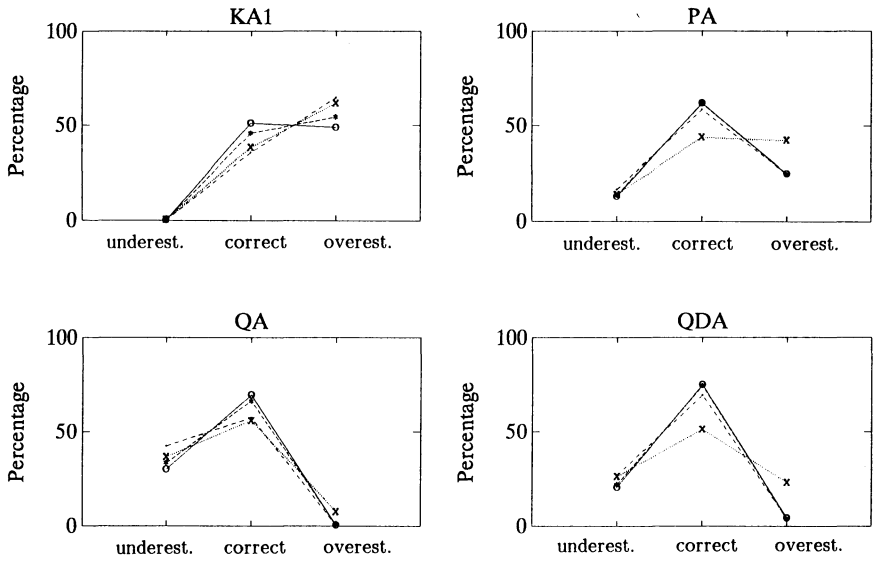


Figure 5.2 *Percentages of times the correct number of components was indicated by each method for four dimension indicators, and the percentages of times over and underestimation occurred*

Answer 1: Because of the superiority of the measure QDA in indicating the correct number of components summed over all conditions of the independent variables and over all methods of analysis, for the remainder of this study the measure QDA was used for dimension indication.

5.6.2 Finding the SCA-method with the best dimension indication

Question 2a: "Which of the SCA-methods has the highest success rate on the preferable dimension indicator?"
Question 2b: "Are there interesting interaction effects between independent variables and method of analysis for the most preferable dimension indicator?"

To get an overview of the size of the main effect of Method and its interactions with one or two between subject factors on the success of

the dimension indicator QDA, the associated sums of squares for the successrate of QDA in the conditions 'samples from one population' and 'samples from two populations', are presented in Table 5.3, where asterisks indicate effects significant at α=.001.

Overall, the four methods differed significantly on the QDA success rate. SCA–P had a significantly higher success rate for the measure QDA (75.3%) than SCA–S (51.0%)and PCA–sep (69.4%), and was also better than SCA–W (74.6%), albeit not significantly. There was a significant interaction effect of Population × Method: The measure QDA had lower success rates in the 'one population' condition than in the 'two populations' condition for SCA–P (71.7% vs. 77.5%), SCA–W (71.4% vs. 76.6%) and PCA–sep (67.8% vs. 70.4%), while for SCA–S the opposite was

Table 5.3: *Sums of Squares for the measure QDA from the ANOVA analysis of samples from one population* ($n = 2160$) *and samples from two populations* ($n = 3600$)

	one population	two populations
Total within subjects	12.00	67.19
Method	1.76*	24.46*
PatternComb × Method	1.40*	6.18*
PhiComb × Method	.10	1.22*
SampSize × Method	.52*	.12
NVar × Method	.67*	10.59*
NFac × Method	.03	5.90*
PatternComb × PhiComb × Method	.65*	1.99*
PatternComb × SampSize × Method	.67*	.32
PatternComb × NVar × Method	.61*	1.54*
PatternComb × NFac × Method	.98*	2.40*
PhiComb × SampSize × Method	.06	.32
PhiComb × NVar × Method	.09	.09
PhiComb × NFac × Method	.11	.72*
SampSize × NVar × Method	.11	.26
SampSize × NFac × Method	.03	.08
NVar × NFac × Method	.05	.72*
Residual within subjects	4.15	10.28

Method = Method used, PatternComb = Used Pattern Combinations, PhiComb = Used Combinations of matrices Φ, *NFac = Number of factors, NVar = Number of variables.* * = *significant at* α = .001.

the case (60.5% vs. 45.4%). Because of this population effect, and because of the incomparability of the patterns used in those two conditions, the results for the QDA measure in both population conditions were inspected.

In Figure 5.3a and 5.3b, the percentages of correct dimension indications by the measure QDA are presented for the SCA-methods and PCA-sep, summed over each condition of each independent variable, for samples from one population and samples from two populations, respectively.

From Figure 5.3a, it can be seen that SCA-W and SCA-P gave similar results for all conditions of the independent variables, while the difference in success of the dimension indicator QDA for SCA-S and PCA-sep compared to SCA-W and SCA-P was relatively small for some conditions of the independent variables and large for other conditions of the independent variables. For the samples from one population, among the independent variables, the greatest differences in the overall success of the dimension indicator QDA are caused by the variable PatternComb. For the conditions '$\mathbf{P}_c\mathbf{P}_c$' and '$\mathbf{P}_e\mathbf{P}_e$' (these are the conditions where the largest amounts of error were present in the data), the dimension indicator QDA performs very poorly, compared to the other conditions of PatternComb. The interaction of PatternComb × Method was the strongest interaction effect, although it explained only 11.7% of the variation (see Table 5.3: Sums of squares of PatternComb × Method is 1.40, see also Figure 5.3a). The effect with the highest sums of squares was Method, explaining 14.7% of the total within subjects variation. The methods SCA-P and SCA-W had the highest percentages of correct dimension indications, SCA-S the lowest, and PCA-sep scored in between. All other sums of squares were considered too small to be relevant.

The results for the samples from two populations (Figure 5.3b) give about the same picture. The success of the dimension indicator for SCA-W and for SCA-P is about the same. This was the case for all conditions of the independent variables, while the difference in success of the dimension indicator QDA for SCA-S and PCA-sep compared to SCA-W and SCA-P was relatively small for some conditions of the independent variables and

large for other conditions of the independent variables. From Table 5.3, we see that the four most relevant effects are the effect of Method and the interactions of Method × PatternComb, Method × NVar, and Method × NFac. The effect of Method is the strongest effect, just as for the samples from one population, explaining 36.4% of the total variation. The interaction of Method × PatternComb can be found, for instance, in small differences between SCA–S and the other methods in PatternComb condition '$\mathbf{P}_c\mathbf{P}_g$' and large differences in '$\mathbf{P}_f\mathbf{P}_g$'. The interaction of Method × NVar comes from the fact that SCA–S performs worse when 24 variables are present than when 12 variables are present, while SCA–W and SCA–P perform better for 24 variables than for 12 variables. The interaction of Method × NFac comes from the fact that the success rate of the measure QDA is about 42% and 36% lower for SCA–W and SCA–P, respectively, when 4 factors are present than when 2 factors are present, while this difference is only about 14% for SCA–S. Together these four effects explain 70.1% of the within subjects variation.

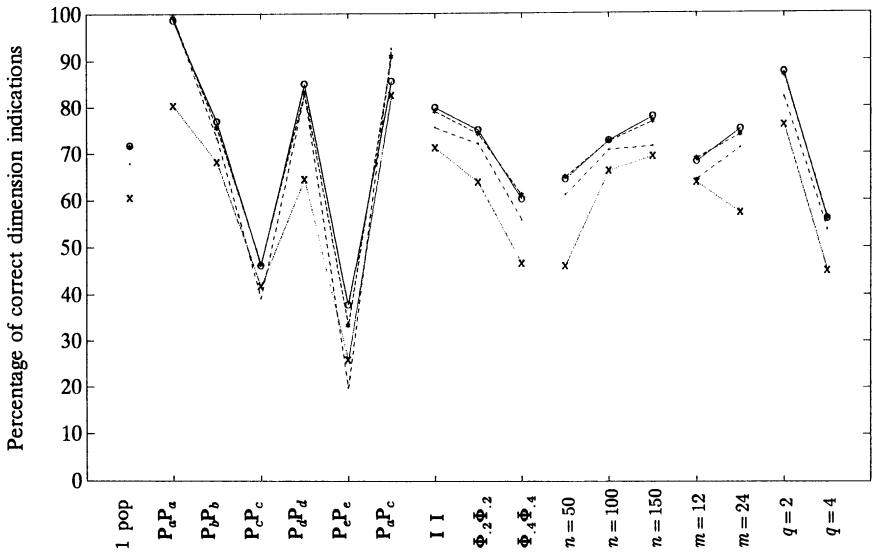


Figure 5.3a *Percentages of correct dimension indications for the three SCA–methods and PCA–sep, averaged over each level of each independent variable, within the samples from one population, using the dimension indicator QDA*

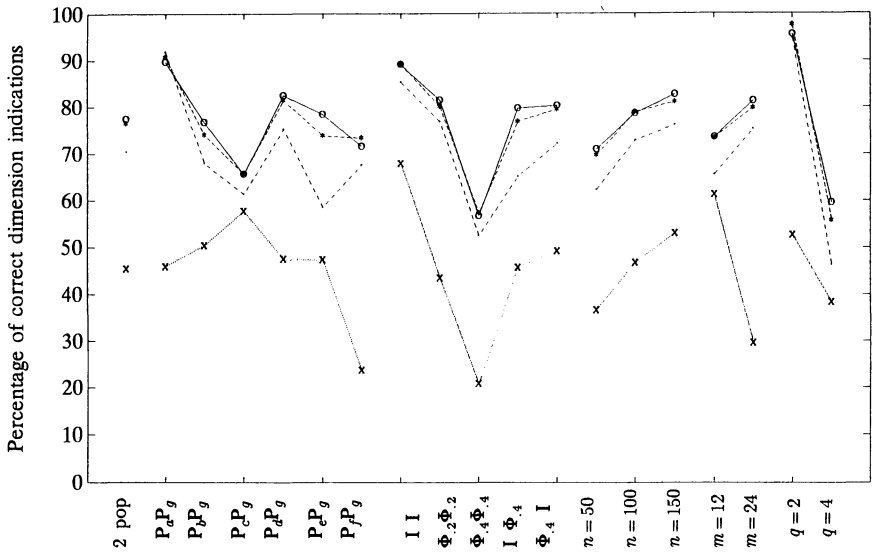
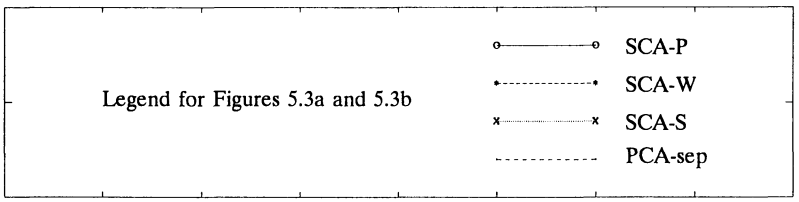


Figure 5.3b *Percentages of correct dimension indications for the three SCA-methods and PCA-sep, averaged over each level of each independent variable, within the samples from two populations, using the dimension indicator QDA*



Concluding, we can say that both in 'samples from one population' and in 'samples from two populations', there was an interesting interaction of PatternComb × Method. The measure QDA indicated the correct number of components much less often for SCA-S than for SCA-W and SCA-P in some of the PatternComb conditions from 'samples from one population' (especially '$\mathbf{P}_a\mathbf{P}_a$' and '$\mathbf{P}_d\mathbf{P}_d$') and in some of the PatternComb conditions from 'samples from two populations' (especially '$\mathbf{P}_a\mathbf{P}_g$' and '$\mathbf{P}_f\mathbf{P}_g$'), while in some of the other PatternComb conditions, success rates for the three SCA-methods hardly differed.

The interaction effect of Population × Method, noted at the

beginning of this Section, can now be seen to be mainly caused by the choices of the different conditions of PatternComb. The two conditions of PatternComb with large amounts of error, used in the condition 'samples from one population', caused the relatively poor performance for SCA–W, SCA–P and PCA–sep, thus giving the impression that the number of components were better indicated for samples from two populations than for samples from one population. However, because the conditions of PatternComb in samples from one population do not have a clear counterpart in the samples from two populations, this conclusion can not be assumed to hold generally. For SCA–S, it can be said that it gave *relatively* better results for the samples from one population than for the samples from two populations, indicating that beside the amount of error, the CM of the true structure matrices is also an important factor in determining the dimensionality correctly.

Answer 2a: From the results, presented above, it was concluded that SCA–W and SCA–P were most successful in indicating the correct number of components, using the measure QDA. These two methods are therefore recommended, when the correct number of components has to be determined solely from the data, that is, when the number of components can not be decided on on theoretical grounds. SCA–P has our special preference, because it takes much less time to execute.

Answer 2b: There was an interesting interaction of PatternComb × Method, caused by the different behavior of the measure QDA for SCA–S, compared to the other two SCA–methods. For data from two populations, there was also an interesting interaction of Method × NVar, again caused by SCA–S. SCA–S gives poorer results when 24 variables are present than when 12 variables are present, while for SCA–W and SCA–P, the opposite is the case.

5.6.3 Finding the SCA–method with the best recovery of factors

Question 3a: "Which of the SCA–methods can be judged as best capable of recovering factors?"
Question 3b: "Are there interesting interaction effects between

independent variables and method of analysis for the recovery of factors?"

In the present experiment, the recovery rate (RR) was computed both using the pattern matrices and using the structure matrices. Whereas the RR was smaller in samples from one population than in samples from two populations for SCA-W and SCA-P (for both the pattern and the structure matrix), the opposite was the case for SCA-S (for both the pattern and the structure matrix). When looking at the RR for the pattern matrices, found with SCA-W and SCA-P, and the structure matrix, found with SCA-S, there was a significant interaction effect of Population × Method on the RR. Therefore, results for the conditions 'samples from one population' and 'samples from two populations' will again be presented separately.

Firstly, the RR's of the pattern matrix and the structure matrix will be presented for each method separately, to show that the use of the pattern matrix for SCA-W and SCA-P and the structure matrix for SCA-S in the final comparison can be justified on empirical grounds, that is, the matrices used had the best overall performance. Secondly, the methods will be compared using the RR measure.

In Figures 5.4a, 5.4b and 5.4c, the RR's of the pattern and the structure matrix are presented for each method separately, summed over each condition of the independent variables, within the one-population data. In Figure 5.5a to 5.5c, the same is done for two-populations data.

For SCA-W and SCA-P, the differences between the RR of the pattern and the structure matrix are small, the pattern matrix being the matrix with the highest overall RR's. For SCA-S, the RR of the structure matrix is the highest for all conditions of the independent variables, within the condition 'samples from one population'. Within the condition 'samples from two populations', the overall RR of the structure matrix is still higher than the RR of the pattern matrix, but for some conditions of the independent variables, the pattern matrix has a higher RR. This is not very surprising, because, when the samples come from different populations, there is not one single structure matrix to be found.

On the basis of these results, it is decided that in the remainder of

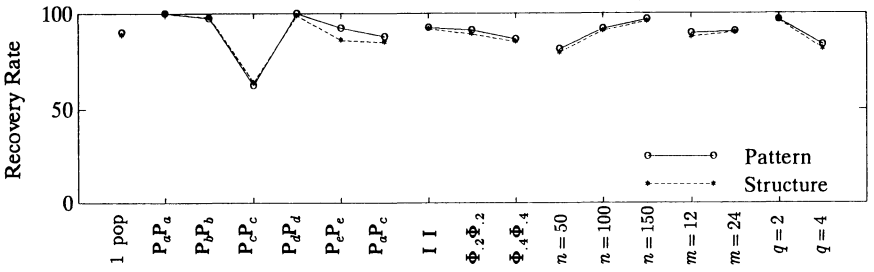


Figure 5.4a *Recovery Rates for the method SCA–W, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from one population*

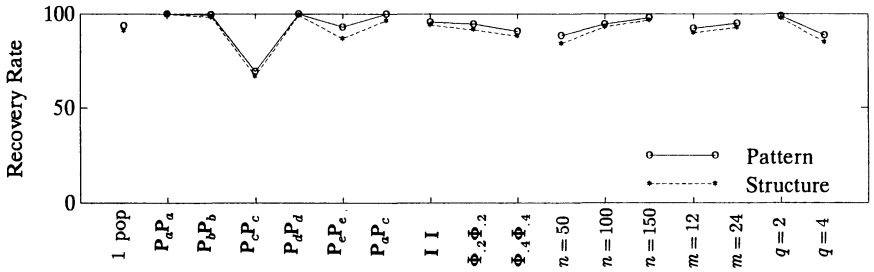


Figure 5.4b *Recovery Rates for the method SCA–P, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from one population*

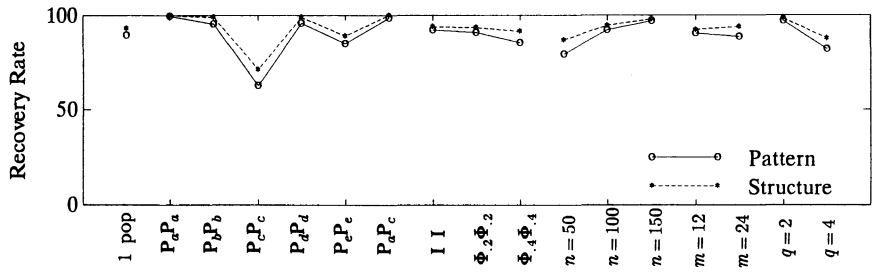


Figure 5.4c *Recovery Rates for the method SCA–S, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from one population*

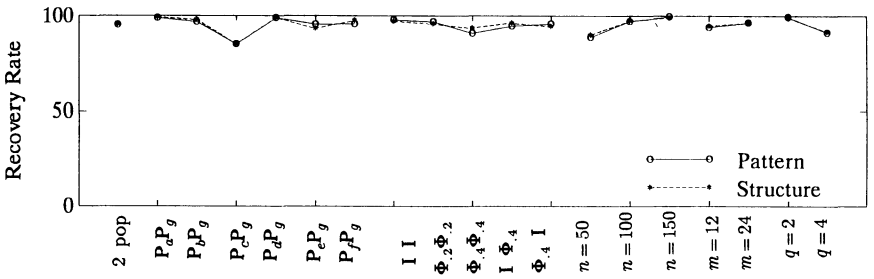


Figure 5.5a *Recovery Rates for the method SCA-W, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from two populations*

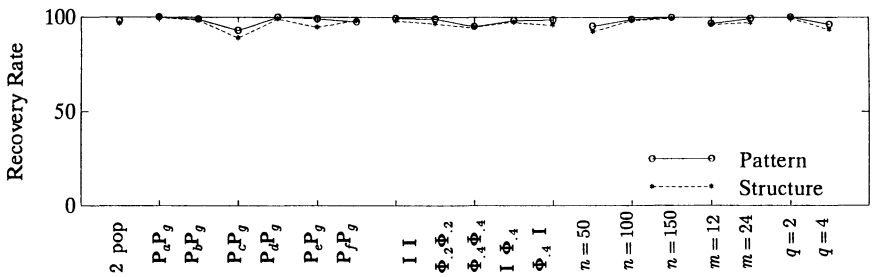


Figure 5.5b *Recovery Rates for the method SCA-P, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from two populations*

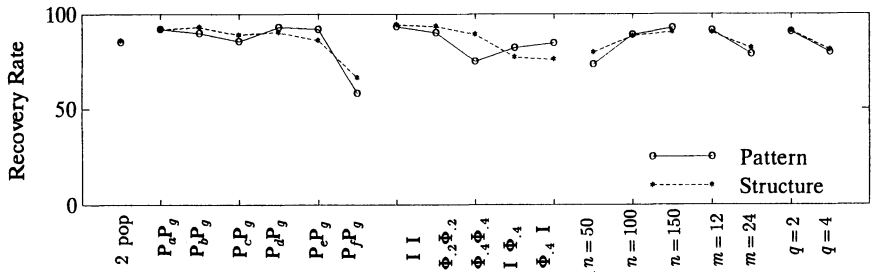


Figure 5.5c *Recovery Rates for the method SCA-S, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from two populations*

this study the RR of the pattern matrix, found with SCA–W, will be compared with the RR of the pattern matrix, found with SCA–P and the RR of the structure matrix, found with SCA–S, as index of factor recovery.

To get an overview of the size of the main effect of Method and its interactions with one or two between subject factors, the associated sums of squares for the success criterion RR in the conditions 'samples from one population' and 'samples from two populations', are presented in Table 5.4, where significant effects at α=.001 are marked by asterisks. The most important effects will be described in Sections 5.6.3.1 and 5.6.3.2.

Table 5.4: *Sums of Squares for the measure RR from the ANOVA analysis of samples from one population* ($n = 2160$) *and samples from two populations* ($n = 3600$)

	one population	two populations
Total within subjects	36.23	147.53
Method	1.67*	27.47*
PatternComb × Method	3.81*	23.03*
PhiComb × Method	.24*	15.32*
SampSize × Method	.86*	1.74*
NVar × Method	.07	6.80*
NFac × Method	.33*	1.76*
PatternComb × PhiComb × Method	.47*	3.06*
PatternComb × SampSize × Method	2.02*	1.59*
PatternComb × NVar × Method	.15	10.40*
PatternComb × NFac × Method	.92*	2.70*
PhiComb × SampSize × Method	.15	1.09*
PhiComb × NVar × Method	.07	.81*
PhiComb × NFac × Method	.07	5.56*
SampSize × NVar × Method	.09	.39*
SampSize × NFac × Method	.07	.62*
NVar × NFac × Method	.11*	.04
Residual within subjects	25.13	45.18

Method = Method used, PatternComb = Used Pattern Combinations, PhiComb = Used Combinations of matrices Φ, *NFac = Number of factors, NVar = Number of variables.* * = *significant at* α = .001.

5.6.3.1 *Samples from one population*

In Figure 5.6, the RR's are presented, summed over each condition of each independent variable, for the pattern matrix, found with SCA-W and SCA-P, and for the structure matrix, found with SCA-S. Overall, a significant Method effect existed (see Table 5.4). The methods SCA-P and SCA-S both had a significantly higher RR (93.6% and 93.1%, respectively) than SCA-W (90.0%). The difference between SCA-P and SCA-S was not significant at α=.001 (p=.015). Summed over each condition of the independent variables SampSize, NVar and NFac separately (these are the independent variables known in practice), SCA-W always had the lowest RR.

The largest effect (see Table 5.4) was the interaction of PatternComb × Method. Looking at the six conditions of the independent variable PatternComb, we see that, when a small amount of error (condition '$\mathbf{P}_a\mathbf{P}_a$', see Appendix A) is added to the factors, all factors are retrieved by the three SCA-methods. With a intermediate amount of error (condition '$\mathbf{P}_b\mathbf{P}_b$'), factors are still almost always retrieved. However, when only weak factors are present (condition '$\mathbf{P}_c\mathbf{P}_c$'), SCA-S, SCA-P, and SCA-W recover only 71.0%, 69.4% and 62.4% of the true factors, respectively, and the differences between methods are considerable.

When half of the factors are strong and half of the factors are intermediate in the population (condition '$\mathbf{P}_d\mathbf{P}_d$'), factors are almost always retrieved. When half of the factors are strong and half of the factors are weak in the population (condition '$\mathbf{P}_e\mathbf{P}_e$'), factors are retrieved in about 90% of the cases, and SCA-S lags about 3% behind with the other two SCA-methods.

Finally, and most interestingly, when factors are strong in the first sample, and weak in the second sample (condition '$\mathbf{P}_a\mathbf{P}_c$') - remember that in this situation both samples are defined as coming from the same population, with as only difference a small amount of measurement error in the first sample and a large amount of measurement error in the second sample - both SCA-P and SCA-S retrieve exactly 100% of the factors, while SCA-W has a RR of only 87.8%. This is the main contribution to the PatternComb × Method interaction. A closer look at these results for

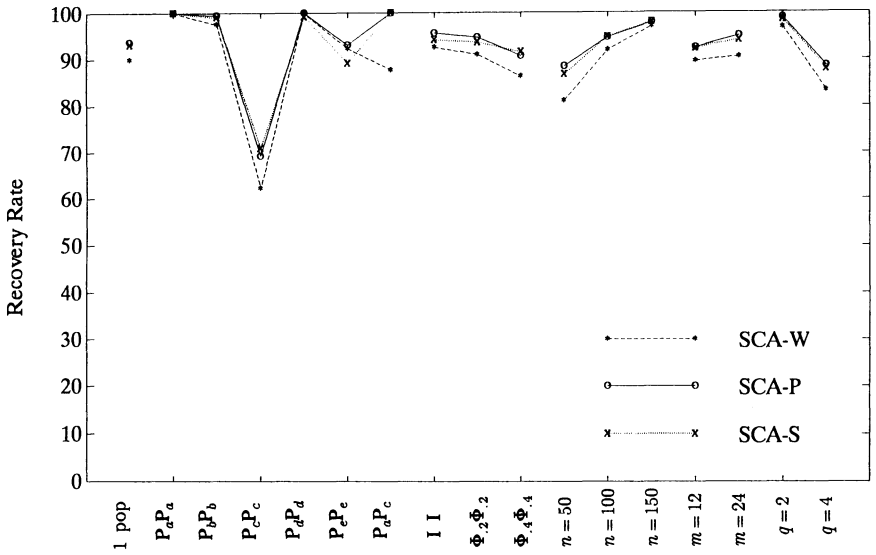


Figure 5.6 *Recovery Rates for the SCA–methods, averaged over each level of each independent variable, within the samples from one population*

SCA–W reveals that SCA–W always retrieves the factors from the first sample with strong factors ($\mathbf{P}_a$), but has a RR of only 75.6% for the factors from the second sample with weak factors ($\mathbf{P}_c$). Separating the '2 factors' from the '4 factors' condition within the condition '$\mathbf{P}_a\mathbf{P}_c$' shows that the RR in the second sample is 88.3% in the '2 factors' condition and 62.8% in the '4 factors' condition. Note that for SCA–S and SCA–P there is hardly any difference between these conditions because the RR is 100%. For PatternComb condition '$\mathbf{P}_c\mathbf{P}_c$', separating the RR's for the '2 factors' and '4 factors' condition gives acceptable RR's of 92.8%, 93.1% and 86.7% for SCA–S, SCA–P, and SCA–W, respectively, in the '2 factors' condition and a troublesome 49.3%, 45.7% and 38.1% for the respective methods in the '4 factors' condition. Together these results explain the significant interaction effect of PatternComb × NFac × Method (see Table 5.4).

There is a significant interaction effect of SampSize × Method.

SCA–W is more vulnerable to small sample sizes than SCA–P and SCA–S. Furthermore, one of the strongest effects is the interaction of PatternComb × SampSize × Method (see Table 5.4). For the conditions '$\mathbf{P}_c\mathbf{P}_c$' and '$\mathbf{P}_a\mathbf{P}_c$' of PatternComb, SCA–W has disproportionately low RR's at small sample size.

In total, the five effects mentioned here explain only a moderate 25.6% of the total variation, while the residual within subjects variation makes up 69.4% of the total variation.

A reason for the low RR's for SCA–W at small sample size can perhaps be found in the fact that, of the SCA–methods, SCA–W is the method that explains the most variance, and has been shown to have a positive sampling bias. This 'greed' for variance to explain may lead the method away from the underlying solution.

5.6.3.2 *Samples from two populations*

In Figure 5.7, the RR's are presented, summed over each condition of each independent variable within the condition 'samples from two populations', for the pattern matrix, found with SCA–W and SCA–P, and for the structure matrix, found with SCA–S. Overall, a highly significant method effect existed, explaining 18.6% of the total variation. SCA–P had a significantly higher RR (98.0%) than both SCA–W (95.1%) and SCA–S (86.2%), as opposed to the results for the '1 population' condition, where both SCA–P and SCA–S had higher RR's than SCA–W. SCA–W had a significantly higher RR than SCA–S. Summed over each condition of the independent variables SampSize, NVar and NFac separately (these are the independent variables that are known in practice), SCA–P always had the highest RR, followed by SCA–W and SCA–S, in that order.

There were strong interactions of PatternComb × Method and PhiComb × Method (see Table 5.4). Results for the six conditions of the independent variable PatternComb will not be discussed in detail, because each of the conditions had a very specific description. For instance, '$\mathbf{P}_c\mathbf{P}_g$' means 'weak factors in one sample' ($\mathbf{P}_c$) and 'factors with variables loading differently within each factor' ($\mathbf{P}_g$). To ease interpretation, it was

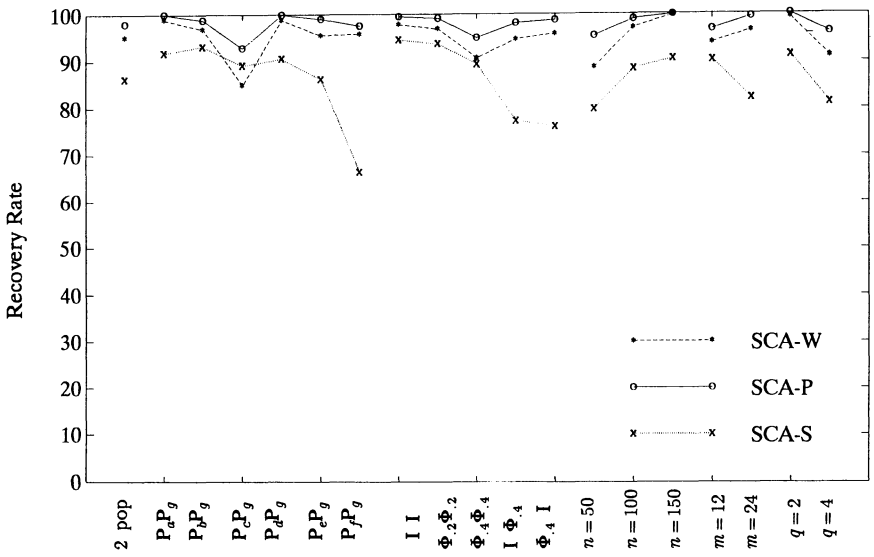


Figure 5.7 *Recovery Rates for the SCA–methods, averaged over each level of each independent variable, within the samples from two populations*

decided to form clusters of conditions combining pattern matrices and matrices Φ used, according to the CM–values of the true structure matrices, given in Table 5.1, and to compare results for these clusters. In order to keep the amount of error, added to the scores, equal over clusters, because this was a factor of great influence on the RR, for each cluster only those conditions were selected for which the mean noise–to–signal ratio used was 1.5. Only the conditions '$\mathbf{P}_b\mathbf{P}_g$', '$\mathbf{P}_e\mathbf{P}_g$' and '$\mathbf{P}_f\mathbf{P}_g$' satisfied this requirement. The fact that the other three conditions of PatternComb are not included here, does not mean that these conditions are dispensable. They were necessary for creating a broad range of possible pattern matrices within the condition 'samples from two populations'. The formed clusters were:

Cluster 1: $CM_{s(tructure)}$ between .88 and .94; $CM_{p(attern)}$ of .94 (n=720);
Cluster 2: CM_s between .80 and .87; CM_p of .94 (n=240);

Cluster 3: CM_s between .70 and .78; CM_p between .78 and .94 (n=720);
Cluster 4: CM_s of .64; CM_p of .78 (n=120).
The average values of the RR for these clusters are presented in Figure 5.8.

From Figure 5.8 it becomes clear that only SCA–S is sensitive to the CM–value of the true–matrices. For Clusters 3 and 4, the RR drops considerably. These results for SCA–S are not very surprising, because the true structure matrices are very different in the two populations in Cluster 4. Thus, it is impossible for SCA–S, which seeks one structure matrix that is the same for all samples, to retrieve both structure matrices. One could expect that SCA–S would better be able to retrieve the true pattern matrices in these situations, because the pattern matrices, found with SCA–S, are allowed to differ across groups. However, the average RR of the pattern matrices is only higher than the average RR of the structure matrix in Clusters 2 and 4, and lower in the other Clusters, as can be seen in Table 5.5. It is remarkable that the RR for SCA–P is as high as it is in Clusters 3 and 4, because the method only produces one pattern matrix. This pattern matrix apparently has high congruence values with both true pattern matrices in Clusters 3 and 4.

These results on differences between RR for pattern and structure led us to inspect the RR of the structure matrix for SCA–W and SCA–P. The RR of the structure matrix is higher than the RR of the pattern matrix in

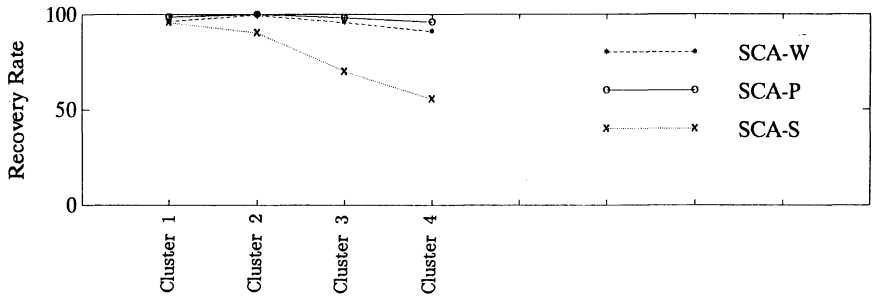


Figure 5.8 *Recovery Rates for the SCA–methods, in four clusters of samples with different (decreasing) CM values for the true structure matrices*

Table 5.5: *Recovery Rates of the pattern and structure matrices, found with the three SCA–methods in the four clusters (in %)*

	SCA–W		SCA–P		SCA–S	
	P	S	P	S	P	S
Cluster 1	96.1	95.8	98.7	96.4	91.5	95.6
Cluster 2	99.5	98.8	100.0	99.3	97.0	90.2
Cluster 3	95.6	96.2	98.1	97.0	66.1	69.8
Cluster 4	90.9	94.9	95.8	97.0	60.1	55.4

Cluster 3 for SCA–W and in Cluster 4 for both methods. So when the differences between the populations become larger, the structure matrix appears to remain closer to the correct factors than the pattern matrix does.

The strong interaction of PhiComb × Method was mainly caused by the low RR's for SCA–S in the conditions '$\mathbf{I}\ \mathbf{\Phi}_{.4}$' and '$\mathbf{\Phi}_{.4}\ \mathbf{I}$'. From Table 5.1 it can be seen that the CM of the true structure matrices is the lowest in the conditions '$\mathbf{I}\ \mathbf{\Phi}_{.4}$' and '$\mathbf{\Phi}_{.4}\ \mathbf{I}$', thus explaining this effect.

There also was a considerable interaction effect of NVar × Method (see Table 5.4). The RR of SCA–S was lower in the condition '24 variables' than in the condition '12 variables', while the RR's of SCA–W and SCA–P were higher in the condition '24 variables' than in the condition '12 variables' (see Figure 5.7). The effect of NVar, just mentioned, was especially different for each method in the PatternComb conditions '$\mathbf{P}_c\mathbf{P}_g$' and '$\mathbf{P}_f\mathbf{P}_g$', causing a strong 3–way interaction of PatternComb × NVar × Method. Worth noting in this respect is the enormous difference in RR for SCA–S within the condition '$\mathbf{P}_f\mathbf{P}_g$', between the conditions '12 variables' and '24 variables'. While in the '12 variables' condition, the RR for the method is 84.0%, in the '24 variables' condition, the RR is 48.6%, a difference of 35.4%. For comparison, within the other conditions of PatternComb, when comparing the '12 variables' with '24 variables' condition, the difference in RR for SCA–S varies between –5.5% and +2.9% (the only positive difference was found in condition '$\mathbf{P}_c\mathbf{P}_g$'). Finally, there was a moderate 3–way interaction of

PhiComb × NFac × Method: PhiComb strongly influences the success of SCA-S. This influence is caused by the low CM of the true structure matrices for two conditions of PhiComb. The negative effect on the RR of SCA-S in these two conditions of PhiComb is stronger when four factors are present than when two factors are present. This influence is again caused by the lower CM of the true structure matrices when four factors are present than when two factors are present (see Table 5.1).

In total, the six effects mentioned here explain 60.0% of the total within subjects variation.

In answer to Question 3a, it is concluded that, when samples are indeed coming from one population (one cannot be certain of this in practice), SCA-P and SCA-S are best capable of recovering the underlying factors, although the differences in success rates between the three SCA-methods are small. When samples are coming from two different populations, SCA-P is still best capable of retrieving the underlying factors, followed by SCA-W and SCA-S, in that order. The capability of SCA-S to retrieve the underlying factors is considerably impaired when the two populations differ markedly. This could be seen as a shortcoming, but according to the author it is actually an asset of the method, because the populations indeed differ, so the method should not be capable of retrieving common factors. It remains a problem, however, how this population difference can be detected in practice (See Section 5.6.5).

In answer to Question 3b, it can be concluded that the independent variable PatternComb has the largest influence on the relative success of each method in recovering underlying factors. It is, furthermore, interesting that for 'samples from two populations', the choice of PhiComb strongly influences the success of SCA-S. This influence is caused by the low CM of the true structure matrices for two conditions of PhiComb. For the 'samples from two populations, the number of variables influences the relative success of each method. While SCA-W and SCA-P have higher recovery rates when 24 variables are present than when 12 variables are present, the opposite is the case for SCA-S. This coincides with the results for the measure QDA.

5.6.4 Finding the SCA-method with the best recovery of correlations between factors

Question 4a: "Which of the SCA–methods can be judged as best capable of recovering the correlations between factors?"
Question 4b: "Are there interesting interaction effects between independent variables and method of analysis for the recovery of correlations between factors?"

The average DFC values (see Section 4.4.2) for samples from one population were .115 for both SCA–W and SCA–P and it was .122 for SCA–S. Note that the DFC values indicate how badly factor correlations are recovered, hence better recovery is indicated by lower values. For the samples from two populations, the average DFC values were .104, .106 and .217 for SCA–W, SCA–P and SCA–S respectively. The different results for samples from one and two populations for SCA–S and the similar results for samples from one and two populations for SCA–W and SCA–P asked for separate analysis of these two conditions.

To get an overview of the size of the main effect of Method and its interactions with one or two between subject factors, the associated sums of squares for the success criterion DFC in the conditions 'samples from one population' and 'samples from two populations', are presented in Table 5.6, where significant effects at α=.001 are again marked by asterisks.

For all conditions of the independent variables within the condition 'samples from one population', the differences between the three SCA–methods were small, although significant, except for NFac. As can be seen from Table 5.6, the total within subjects variation was very small in the condition 'samples from one population', as compared with the condition 'samples from two populations'. Even when correcting for the larger number of cases in the latter condition, the variation in the condition 'samples from two populations' was still 17 times the variation in the condition 'samples from one population'. The largest effect in the condition 'samples from one population' was PhiComb × Method. The mean

Table 5.6: *Sums of Squares for the measure DFC from the ANOVA analysis of samples from one population* ($n = 2160$) *and samples from two populations* ($n = 3600$)

	one population	two populations
Total within subjects	2.44	70.41
Method	.07*	30.15*
PatternComb × Method	.08*	.61*
PhiComb × Method	.40*	26.28*
SampSize × Method	.17*	.07*
NVar × Method	.09*	.26*
NFac × Method	.00	.20*
PatternComb × PhiComb × Method	.20*	1.94*
PatternComb × SampSize × Method	.04*	.12*
PatternComb × NVar × Method	.01	.20*
PatternComb × NFac × Method	.02*	.10*
PhiComb × SampSize × Method	.03*	.54*
PhiComb × NVar × Method	.01	.18*
PhiComb × NFac × Method	.02*	.12*
SampSize × NVar × Method	.05*	.00
SampSize × NFac × Method	.00	.02
NVar × NFac × Method	.00	.17*
Residual within subjects	1.25	9.43

Method = Method used, PatternComb = Used Pattern Combinations, PhiComb = Used Combinations of matrices Φ, *NFac = Number of factors, NVar = Number of variables.* * = *significant at* α = .001.

values of the DFC measure for SCA-W, SCA-P and SCA-S, were respectively .08, .09, and .11 in PhiComb condition '**I I**', .11, .11, and .12 in PhiComb condition '$\Phi_{.2}\Phi_{.2}$', and .16, .15, and .14 in PhiComb condition '$\Phi_{.4}\Phi_{.4}$', illustrating that differences in differences between the methods were indeed small. Therefore, the differences, found in the condition 'samples from one population', were deemed not interesting.

Within the condition 'samples from two populations' there was a sizable Method effect. The mean DFC value was larger for SCA-S than for SCA-W and SCA-P. By far the strongest interaction was of PhiComb × Method. In Figure 5.9, DFC values for the five conditions of the independent variable PhiComb, within the condition 'samples from two populations', are presented for each method.

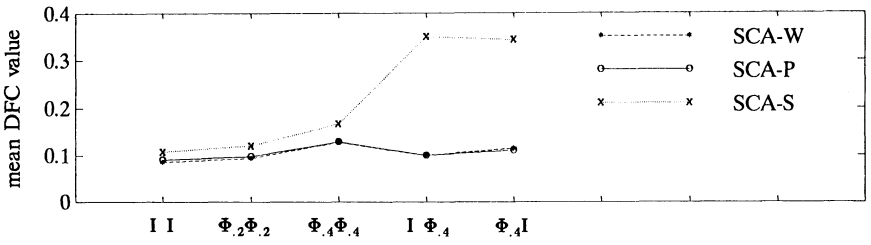


Figure 5.9 *Mean DFC values for the SCA-methods, averaged over the five levels of the independent variable PhiComb, within the samples from two populations*

From Figure 5.9, it can be seen that in the conditions '**I** $\Phi_{.4}$' and '$\Phi_{.4}$ **I**' SCA–S has a much larger DFC value than SCA–W and SCA–P. Recall that for these two conditions of PhiComb the CM of the true structure matrices was the lowest, and the RR of the structure matrix by SCA–S was also the lowest. Together, the effects of Method and the interaction of PhiComb × Method explain 80.1% of the total variation between subjects.

Answers 4a and 4b: For the samples from one population, there were no large differences between the SCA–methods in recovering the correlations between the factors. When samples come from two different populations, SCA–W and SCA–P give correlations closer to the correct values than SCA–S. This difference is mainly found in the condition where the samples from two populations are based on different factor correlation matrices.

5.6.5 Discerning the 'samples from one population' from the 'samples from two populations' condition

Question 5: "Can samples from one population be distinguished from samples from two populations?"

A discriminant analysis was performed to see to what extent we can

correctly classify each data set (consisting of two samples) in the category 'samples from one population' or in the category 'samples from two populations'. A subset of data sets was selected for inclusion in the analysis. The selection was made on the basis of two requirements. First, the overall amount of error in the data sets had to be the same in all data sets, because otherwise the (in practice unobservable) amount of error in the data could be the strongest aspect influencing the discriminant function and the desired discriminant function could be blurred. Secondly, the selected PatternComb conditions on which the samples from two populations were based had to be 'considerably' different, that is, only those data sets were selected for which the CM of the true pattern matrices was lower than .85 (see Table 5.1). On the basis of these two requirements, the 1080 data sets based on the PatternComb conditions '$\mathbf{P}_b\mathbf{P}_b$', '$\mathbf{P}_e\mathbf{P}_e$' and '$\mathbf{P}_a\mathbf{P}_c$' were selected to represent samples from one population, and the 600 data sets based on the PatternComb condition '$\mathbf{P}_f\mathbf{P}_g$' were selected to represent samples from two populations.

In the discriminant analysis, the following seven predictor variables were used: SampSize, NVar, NFac and the percentages of variance explained by the three SCA–methods and PCA–sep ($EV_{SCA\text{-}W}$, $EV_{SCA\text{-}P}$, $EV_{SCA\text{-}S}$ and $EV_{PCA\text{-}sep}$). These seven predictor variables are all the variables that are observed (or can be observed) in practice.

In a first analysis, all of the predictors mentioned above were included. The overall percentage of correct classifications was 92.2%. The percentage of correct classifications was 95.8% for samples from one population (PatternComb conditions '$\mathbf{P}_b\mathbf{P}_b$', '$\mathbf{P}_e\mathbf{P}_e$' and '$\mathbf{P}_a\mathbf{P}_c$') and 85.7% for samples from two populations (PatternComb condition '$\mathbf{P}_f\mathbf{P}_g$'). The predictor variable 'percentage of explained variance of SCA–P' did not pass the tolerance test and was excluded from the discriminant function. The tolerance is a measure of the degree of linear association between the independent variables, implemented in SPSS (Norušis, 1990): "For the i^{th} independent variable, it is [calculated as] $1–R_i^2$, where R_i^2 is the squared multiple correlation coefficient when the i^{th} independent variable is considered the dependent variable and the regression equation

between it and the other independent variables is calculated. Small values for the tolerance indicate that the i^{th} independent variable is almost a linear combination of the other independent variables. Variables with small tolerances (by default, less than .001) are not permitted to enter the analysis."

The predictor variable 'percentage of explained variance of SCA-W' had a very small standardized coefficient (-.15), compared to that for the explained variances of PCA-sep (-3.00) and SCA-S (4.04). Because of this, a second discriminant analysis was done, leaving out the explained variances of SCA-P and SCA-W as predictor variables. This resulted in a discriminant function that correctly classified the cases about just as often (overall percentage of correct classifications was 92.3%; 95.8% for samples from one population and 85.8% for samples from two populations). Finally, a third discriminant analysis was done to see whether inclusion of the predictor variables SampSize, NVar and NFac was necessary, so only the explained variance of PCA-sep and SCA-S were used as predictor variables. The percentages of correct classifications was now 86.8% for the samples from one population and 77.2% for the samples from two populations (overall 83.3%). The loss in correct classifications in the third analysis was considered too large, and therefore the discriminant function from the second analysis was considered most useful for classifying data sets as coming from one population or from two 'considerably' different populations.

To crossvalidate the results from the discriminant analysis, the selection of cases used for the discriminant analysis (PatternComb conditions '$\mathbf{P}_b\mathbf{P}_b$', '$\mathbf{P}_e\mathbf{P}_e$', '$\mathbf{P}_a\mathbf{P}_c$' and '$\mathbf{P}_f\mathbf{P}_g$') was split in two. That is, the ten replications within each condition were evenly divided over the two halves. On each of the halves, denoted as A and B, another discriminant analysis was performed and the two discriminant functions thus found, denoted as D_A and D_B, were put to the test on the other half. The percentages of correct classifications using D_A on A were 97.0% for samples from one population and 84.7% for samples from two populations. Using D_B on A gave correct classifications of 98.0% and 83.3%, respectively. (Specifically, in total 779 cases were correctly classified

in the crossvalidation of D_B on A, against 778 when using D_A on A. This perhaps counterintuitive improvement is not disturbing, because the optimized discriminant function does not necessarily gives the best classification; it only does so when the distribution of the groups discriminant scores is normal.) The percentages of correct classifications using D_B on B were 94.8% for samples from one population and 85.0% for samples from two populations. Using D_A on B gave correct classifications of 90.0% and 89.3%, respectively. The percentages of correct classifications show good crossvalidation of the results. The coefficients in the two discriminant functions were very similar, differing .02 at most, except for the predictor variable NFac, where the unstandardized coefficient was −.36 in D_A and −.26 in D_B. This predictor variable also had the smallest standardized discriminant function coefficient. From this crossvalidation, it was concluded that the discriminant weights were very stable, except for the weights for predictor variable NFac. For this reason, it was decided do a final discriminant analysis, in which that predictor variable was left out. To further simplify the thus found discriminant function, a simple version of it was tested, in which only the first three decimals from the complete discriminant function were used. Denoting the percentage of explained variance by EV, the simple discriminant function can be written as

$$d(EV_S,EV_A,NVar,SS) = .469EV_S-.431EV_A+.136NVar-.016SS+.453, \qquad (5.1)$$

in which $EV_S=EV_{SCA\text{-}S}$, $EV_A=EV_{PCA\text{-}sep}$ and SS=SampSize. When d>0, the discriminant function indicates that the samples are *probably* from one population. When d<0, it indicates that the samples are *probably* from two different populations.

As a second validation, the complete discriminant function (i.e., function (5.1) with 7 decimals instead of three) and the simple discriminant function were used on all data sets, categorized with respect to the twelve different PatternComb conditions used in this experiment. The percentages of correct classifications for each PatternComb condition are given in Table 5.7.

From Table 5.7, it can be seen that there are only small differences

Table 5.7: *Percentages of correct classifications for the two discriminant functions for each condition of PatternComb*

	complete function	*simple function*
One–population data		
$\mathbf{P}_a\mathbf{P}_a$	95.6	95.6
$\mathbf{P}_b\mathbf{P}_b$	96.9*	96.9*
$\mathbf{P}_c\mathbf{P}_c$	79.4	82.2
$\mathbf{P}_d\mathbf{P}_d$	97.8	97.8
$\mathbf{P}_e\mathbf{P}_e$	98.6*	98.6*
$\mathbf{P}_a\mathbf{P}_c$	95.0*	95.3*
Two–or–more–populations data		
$\mathbf{P}_a\mathbf{P}_g$	55.2	52.0
$\mathbf{P}_b\mathbf{P}_g$	27.7	24.2
$\mathbf{P}_c\mathbf{P}_g$	18.7	15.8
$\mathbf{P}_d\mathbf{P}_g$	43.3	40.2
$\mathbf{P}_e\mathbf{P}_g$	26.8	21.5
$\mathbf{P}_f\mathbf{P}_g$	80.5*	78.3*

* *indicates conditions used in the discriminant analysis*

between the two discriminant functions, thus suggesting that the simple function can be used in the sequel. However, some at first sight disturbing results were found. In the crossvalidation cases, the discriminant function only gives good results for the PatternComb conditions '$\mathbf{P}_a\mathbf{P}_a$' and '$\mathbf{P}_d\mathbf{P}_d$'. For all other PatternComb conditions, the discriminant function classifies the cases wrongly for more than 50% of the data sets. The question to ask at this point is: "How serious are these misclassifications?"

The fact that samples from the PatternComb condition '$\mathbf{P}_c\mathbf{P}_c$' are classified as coming from two populations in 53% of the cases is not disturbing, because these samples are very poor anyway (they contain a very large amount of error), and thus a researcher is obliged to proceed with caution. The fact that samples based on the PatternComb conditions '$\mathbf{P}_a\mathbf{P}_g$' to '$\mathbf{P}_e\mathbf{P}_g$' are often classified as coming from one population is also not very disturbing, because from the CM of the true pattern matrices (see Table 5.1) it can be seen that the samples come from

different, but not 'considerably' different populations, and thus, in practice, it is acceptable to view the samples as coming from one population. Finally, with the present discriminant function an easy rule of thumb can be given for discerning samples from one population with samples from two populations. If the discriminant value indicates that the samples come from two different populations, it is very likely that this is indeed the case.

Answer 5: By using discriminant function (5.1), it can be determined reasonably reliably whether or not samples come from two populations with distinctly different underlying patterns.

5.6.6 Recovery of factors for specific cases of samples from one population

To investigate the ability of the SCA-methods to cope with different variations in strength of factors, several separate comparisons were made.

Question 6a: "Which SCA-method gives the highest RR, when comparing samples taken from one population, with only strong factors (condition '$\mathbf{P}_a\mathbf{P}_a$'), only intermediate factors (condition '$\mathbf{P}_b\mathbf{P}_b$'), and only weak factors (condition '$\mathbf{P}_c\mathbf{P}_c$')?"

In Table 5.8, the RR's for the SCA-methods are given for four conditions of PatternComb (within each PatternComb condition RR's are averaged over all other independent variables), describing samples from one population. The first three conditions ('$\mathbf{P}_a\mathbf{P}_a$', '$\mathbf{P}_b\mathbf{P}_b$' and '$\mathbf{P}_c\mathbf{P}_c$') are also depicted in Figure 5.6. For the condition in which factors of different strength were present in the two samples ('$\mathbf{P}_a\mathbf{P}_c$'), RR's for the factors of different strength are given separately.

There was a significant interaction effect between PatternComb and Method (see Figure 5.6). Comparing two methods at a time, it appeared that SCA-W showed a significantly larger drop in RR when going from pattern '$\mathbf{P}_a\mathbf{P}_a$' to '$\mathbf{P}_c\mathbf{P}_c$', than both SCA-S and SCA-P. The difference

Table 5.8: *RR's for the SCA-methods for the conditions of PatternComb describing samples from one population, with RR's separated for factors of different strength*

	SCA-W		SCA-P		SCA-S	
$\mathbf{P}_a\mathbf{P}_a$ $(=s)$	1.00		1.00		1.00	
$\mathbf{P}_b\mathbf{P}_b$ $(=i)$	.98		.99		.99	
$\mathbf{P}_c\mathbf{P}_c$ $(=w)$	.62		.69		.71	
Fact.str.	*s*	*w*	*s*	*w*	*s*	*w*
$\mathbf{P}_a\mathbf{P}_c$	1.00	.76	1.00	1.00	1.00	1.00

s=strong (loading=.89), i=intermediate strength (.55), w=weak (.37)

between SCA-S and SCA-P was not significant at α=.001 (p=.048).

Answer 6a: When large amounts of error are present, SCA-P and SCA-S give the best results on the RR measure. When small or intermediate amounts of error are present, all methods have almost complete recovery of factors.

Question 6b: "Are, for the SCA-methods, intermediate factors and weak factors blurred by strong factors, or vice versa?"

For this question, the influence of factors of different strengths within a sample was investigated. To clarify the comparison, keep the following situation in mind. One has scores on a set of 24 variables, pertaining to 4 factors: 2 strong factors and 2 intermediate or weak factors (as in $\mathbf{P}_d$ and $\mathbf{P}_e$). Each factor is defined by nonzero loadings on 6 variables. If one would analyze the 12 variables making up the intermediate or weak factors separately (as in data based on $\mathbf{P}_b$ and $\mathbf{P}_c$), one would like to see the same results for these factors as when analyzing the complete data set with 24 variables. To see whether this requirement is fulfilled by the SCA-methods, the '12 variables and 2 factors' conditions within the conditions '$\mathbf{P}_b\mathbf{P}_b$' and '$\mathbf{P}_c\mathbf{P}_c$' were compared with the '24 variables and 4 factors' conditions within the conditions '$\mathbf{P}_d\mathbf{P}_d$' and '$\mathbf{P}_e\mathbf{P}_e$', respectively (see Table 5.9). Also comparisons of

condition '$\mathbf{P}_a\mathbf{P}_a$' and the conditions '$\mathbf{P}_d\mathbf{P}_d$' and '$\mathbf{P}_e\mathbf{P}_e$' were made. In these comparisons the variable-factor ratio is held constant.

Comparing the RR of the strong factors in condition '$\mathbf{P}_a\mathbf{P}_a$' with '$\mathbf{P}_d\mathbf{P}_d$' and '$\mathbf{P}_e\mathbf{P}_e$', we see that the RR for the strong factors drops for SCA-W and SCA-S, when weak factors are added to the analysis. SCA-P is not troubled by this effect. Comparing the RR of the factors of intermediate strength in condition '$\mathbf{P}_b\mathbf{P}_b$' with '$\mathbf{P}_d\mathbf{P}_d$', we see no effects, because recovery of the factors is 100%, except for a few factors not recovered by SCA-S (10 out of 360). Comparing the RR of the weak factors in condition '$\mathbf{P}_c\mathbf{P}_c$' with '$\mathbf{P}_e\mathbf{P}_e$', we see that the RR for SCA-P drops from .88 to .77, while it drops very little for both SCA-W and SCA-S. So, in this condition, SCA-P seems to favor the recovery of the strong factors at the cost of the recovery of the weak factors, and SCA-W and SCA-S seem to favor the recovery of the weak factors at the cost of the recovery of the strong factors.

Answer 6b: In the condition where the variable-factor ratio is held equal, for SCA-W and SCA-S, it was found that the presence of weak factors, within each sample, decreases the recovery of strong factors. For SCA-P, the presence of strong factors, within each sample, decreases the recovery of weak factors.

Table 5.9: *RR's for the SCA-methods for the* 12 *variables/*2 *factors conditions of levels* '$\mathbf{P}_a\mathbf{P}_a$', '$\mathbf{P}_b\mathbf{P}_b$' *and* '$\mathbf{P}_c\mathbf{P}_c$', *and for the* 24 *variables/*4 *factors conditions of levels* '$\mathbf{P}_d\mathbf{P}_d$' *and* '$\mathbf{P}_e\mathbf{P}_e$' (*number of factors per cell is* $n = 360$)

	SCA-W		SCA-P		SCA-S	
$\mathbf{P}_a\mathbf{P}_a$ $(=s)$	1.00		1.00		1.00	
$\mathbf{P}_b\mathbf{P}_b$ $(=i)$	1.00		1.00		1.00	
$\mathbf{P}_c\mathbf{P}_c$ $(=w)$	.84		.88		.90	
Fact.str.	*s*	*i*	*s*	*i*	*s*	*i*
$\mathbf{P}_d\mathbf{P}_d$	1.00	1.00	1.00	1.00	.98	.97
Fact.str.	*s*	*w*	*s*	*w*	*s*	*w*
$\mathbf{P}_e\mathbf{P}_e$	.90	.83	1.00	.77	.81	.87

s=strong (loading=.89), i=intermediate strength (.55), w=weak (.37)

Question 6c: "Is the RR for the three SCA–methods affected by the existence of small measurement error in one sample and large measurement error in another sample (condition '$\mathbf{P}_a\mathbf{P}_c$'), and which SCA–method performs best under this condition?"

We return to Table 5.6 one more time to compare the RR for the SCA–methods in the conditions '$\mathbf{P}_a\mathbf{P}_a$' and '$\mathbf{P}_c\mathbf{P}_c$' with '$\mathbf{P}_a\mathbf{P}_c$', to see whether RR's were affected by the existence of small measurement error in one sample and large measurement error in another sample. From Table 5.8, we see that the strong factors were always recovered. Furthermore, for all SCA–methods the recovery of the weak factors is enhanced when one of the samples contains strong factors (scores with little measurement error). SCA–W, however, benefits only slightly from the presence of a sample with strong factors.

Answer 6c: The methods SCA–P and SCA–S are fully capable of retrieving weak factors in one sample, when the same factors in the other sample are strong. SCA–W does not have this capability.

5.7 CONCLUSION

Of the four dimension indicators tested, the measure QDA gave the best overall results for indicating the correct number of components for all methods of analysis. For this reason, the measure QDA will be used as the success criterion for dimension indication in the remainder of this study.

SCA–P gives the best results for all dimension indicators. The measure QDA is the dimension indicator that can best be used when applying SCA–W, SCA–P and PCA–sep, and the measure QA is the dimension indicator that can best be used when applying SCA–S. Surprisingly, differences between the dimension indicators are not much influenced by sample size, although all dimension indicators have higher success rates for larger sample sizes. The effect of the number of factors present is

much stronger. Four components are much harder to be correctly indicated by the measures PA, QA and QDA, while the measure KA1 has more trouble correctly indicating two components. When four factors are present, all four dimension indicators perform about equally well for all methods, although the measure QDA still has the highest success rate.

Using the measure QDA, SCA-W and SCA-P (including the rotations, as described in Chapter 2) were most successful in indicating the correct number of components. SCA-P is the method recommended, when the correct number of components has to be determined solely from the data, that is, when the number of components can not be decided on on theoretical grounds, because this method takes the shortest time to execute.

When samples are coming from one population, SCA-P and SCA-S are somewhat better capable of recovering the underlying factors than SCA-W. This method especially falls behind when only weak factors are present and when there are weak factors in one sample, when the same factors in the other sample are strong. SCA-W fails to recover factors in situations where, compared to the other SCA-methods, it can gain a relatively large percentage of variance. These situations are at small sample sizes (see Section 5.6.3.1) and when there are weak components present (large amounts of error). SCA-W seems to show the property of encompassing a relatively large portion of 'error' in its components. It must be noted, however, that mostly the differences in success rates between the three SCA-methods are small. The lower recovery rate of SCA-W in the conditions mentioned above is especially apparent at small sample size (n=50). For SCA-P, the presence of strong factors within each sample decreases the recovery of weak factors. For SCA-W and SCA-S, the presence of weak factors within each sample decreases the recovery of strong factors. When samples are coming from two different populations, SCA-P is best capable of retrieving the underlying factors, followed by SCA-W and SCA-S, in that order.

The results on retrieving the correlations between the factors reveal that for samples from one population, all three SCA-methods perform about equally good. When samples come from two 'considerably' different populations, SCA-W and SCA-P give correlations closer to the

correct values than SCA–S.

Overall, SCA–P has the best results on dimension indication and on retrieving factors and factor correlations and is therefore recommended. SCA–W performs somewhat more poorly, and SCA–S performs about just as well as SCA–P when the samples come from one population, and performs worse than SCA–P when the samples come from two populations.

If it is desired, when doing simultaneous components analyses, to determine reasonably reliably whether samples come from two populations with distinctly different underlying patterns or not, we would advise to use discriminant function (5.1). It is, however, advised to use the discriminant function only when the number of variables and the sample size used are within the range of the values used in the determination of the discriminant function. Furthermore, it might be wise to use both SCA–P and SCA–S, and see whether the two methods lead to different interpretations of the components found. If so, this is an additional indication that there are considerable differences between the populations from which the samples were drawn, because in this case the solutions from SCA–S tend to differ considerably from the solutions from SCA–P (where the latter still adequately recover underlying factors, whereas the former to a large extent fail to do so).

CHAPTER 6

EXPERIMENT 2, TESTING THE METHOD SIFASP-ML

6.1 INTRODUCTION

In the second experiment, conducted in this study, the objective was to see how well SIFASP–ML held up under various circumstances and how SIFASP–ML compared with the three SCA–methods. For this, a selection of the data manipulations, as used in Experiment 1, was chosen. Raw data were constructed as described in Section 3.2.1, but because Jöreskog (1971, pp. 417–418) argued that only scores standardized over groups may be used for SIFASP, SIFASP–ML was used with such standardized scores. The covariances obtained from the thus standardized scores were analyzed by SIFASP–ML. For comparison, the simulated scores were also standardized *within* each sample, thus arriving at matrices with correlations. These correlations were also analyzed by SIFASP–ML. Note that all correlation matrices that were analyzed by SIFASP–ML have also been analyzed by the SCA–methods in Experiment 1. The results for SIFASP–ML were compared with the results for SCA–P for the same conditions.

6.2 MANIPULATIONS OF THE DATA

The following values were chosen for the independent variables. In the simulated data sets there were two groups (p=2), 12 or 24 variables (=m), 2 or 4 factors (=q) and group sizes of n=50, n=100 or n=150. The true pattern matrices and matrices Φ used in Experiment 2 are a selection of the matrices used in Experiment 1, as given in Appendix A. A selection was made for the primary reason that running the program was rather laborious. In our view, (the combinations of) the true pattern matrices used still form a reasonable representation of the different situations

one might expect to encounter in the "real life" research domain. The conditions of PatternComb and PhiComb, used to simulate one–population data, were '$\mathbf{P}_a\mathbf{P}_a$' and '$\mathbf{P}_e\mathbf{P}_e$', and '**I I**' and '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$', respectively. The conditions of PatternComb and PhiComb, used to simulate two–populations data, were '$\mathbf{P}_a\mathbf{P}_g$' and '$\mathbf{P}_f\mathbf{P}_g$', and '**I I**' and '**I** $\mathbf{\Phi}_{.4}$', respectively. For the value of the CM of the true pattern and structure matrices for each condition, we refer to Table 5.1.

In total, there were (2×2×3×4=) 48 one–population data conditions and 48 two–populations data conditions. In each condition, ten replications of simulated data were used.

6.3 DEPENDENT VARIABLES

For each data set of two groups, six models were fitted in which 1, 2, 3, 4, 5 and 6 factors were specified, respectively. The identification constraints (see Section 2.3.2) employed with these models are given in Appendix B. For each of the six models (if the estimation procedure converged), the values of the fit indices were obtained.

According to the description of the use of the dimension indicators, given in Section 4.3.6, it was recorded for each data set whether under–, correct or overestimation of the number of factors occurred. For each solution with the correct dimension, the measures RR and DFC were calculated (see Sections 4.4.1 and 4.4.2).

6.4 ANALYSIS

All the measures presented were analyzed with repeated measures analyses of variance. Sums of squares of main and interaction effects were calculated to give an impression of the size of the effects, and the associated averaged univariate F tests were employed for significance testing. The same adjustments and checks for violations of assumptions were used as in Experiment 1.

The analyses were performed on the data sets for one-population data and two-or-more-populations data, separately. The between subjects variables in the analyses were: PatternComb (2 conditions; note that for one-population data and two-or-more-populations data there were 2 different conditions of PatternComb), PhiComb (2 conditions for one-population data and 2 conditions for two-or-more-populations data), SampSize (3 conditions; n=50, 100, 150), NVar (2 conditions; 12 and 24 variables), and NFac (2 conditions; 2 and 4 factors). The variable Method was specified as a within subjects factor (2 conditions; the best of the methods SIFASP-ML applied to covariances and SIFASP-ML applied to correlations, and SCA-P (the best of the SCA-methods)) in a repeated measures design, because each method analyzed the same data sets. In each of the separate analyses, only first and second order interactions among between factors were included in the analysis.

For the analysis of the results for the dimension indicators, observations concerning correct or incorrect indication of the number of factors were aggregated over the ten replications within each condition to obtain a reasonably continuous variable: The success rate of indicating the correct number of factors over ten replications. This meant that for each dimension indicator for each method, for each condition a value between 0 and 1.0 was obtained. The dimension indicator used was specified as a within subjects factor (4 levels; SCDT, RMSEA, ECVI and COMBI) in a repeated measures design. The best of the fit indices was compared with the best of the dimension indicators (QDA) for the SCA-methods, again in a repeated measures design, with the dimension indicator specified as a within subjects factor with two levels.

For the success criterion RR, the RR of the overall best of the two SIFASP-ML methods was analyzed together with the RR for SCA-P in a repeated measures design. For the analysis of the success criterion DFC the same was done.

6.5 QUESTIONS TO BE ANSWERED BY THIS EXPERIMENT

The present experiment was conducted to get answers as to which fit index can best be used for determining the number of factors present in a data set, what factors play a role in the success and failure in the recovery of factors by SIFASP–ML, and how SIFASP–ML compares with SCA–P, the best of the SCA–methods. The question numbers refer to the subsections of Section 6.6, in which answers to these questions will be given. Summarizing conclusions from the results of this experiment will be drawn in Section 6.7.

1a.) Which of the fit indices SCDT, RMSEA, ECVI and COMBI can best be used to determine the number of factors?
1b.) Which of the methods SIFASP–ML and SCA–P is best in indicating the correct dimension?
2a.) Is the highest RR found when analyzing covariances (scores standardized over groups) or when analyzing correlations?
2b.) Which of the methods SIFASP–ML and SCA–P has the highest RR?
2c.) What, if any, are the most interesting interaction effects between independent variables and method of analysis for the recovery of factors?
3a.) Is the smallest DFC found when analyzing covariances (scores standardized over groups) or when analyzing correlations?
3b.) Which of the methods SIFASP–ML and SCA–P has the lowest DFC?
3c.) What, if any, are the most interesting interaction effects between independent variables and method of analysis for the recovery of correlations between factors?
4.) How often do the solutions for the correct dimension have bad fit? For this purpose, the rules of thumb, mentioned in Section 4.3.5, were used.

6.6 RESULTS

In the present experiment, a significance level α=.001 was used.

Whenever differences between dependent variables or methods are reported, or when interaction effects are said to occur, effects were significant at α=.001. Whenever interesting results with a p-value larger than .001 were encountered, the p-value is reported. In the following Sections, each of the six main questions, asked at the end of Section 6.5, will be answered separately. At the start of each section, the question(s) will be repeated.

6.6.1 Finding the preferable dimension indicator

Question 1a: "Which of the fit indices SCDT, RMSEA, ECVI and COMBI can best be used to determine the number of factors?"

To answer this question, the four dimension indicators were compared for SIFASP-ML applied to covariances (from now on denoted as SIFASP-ML-C) and SIFASP-ML applied to correlations (from now on denoted as SIFASP-ML-R) separately. In Figures 6.1a through 6.1d, the percentage of times the model with the correct number of factors was indicated by the four fit indices is presented for both methods separately. The percentages are given, summed over all conditions taken together, and summed over each condition of each independent variable separately. For calculating the success rate of the dimension indicators, two subsets of solutions were used. Firstly, the success rate of the dimension indicators was calculated, leaving out all nonconverged and nonadmissible solutions (Figures 6.1a and 6.1b). Secondly, the success rate of the dimension indicators was calculated, when only the nonconverged solutions were left out (Figures 6.1c and 6.1d). It should be noted that leaving out nonconverged and/or nonadmissible solutions may just as well increase the success rate (when solutions of wrong dimensions are thus left out) as decrease the solutions (when solutions of the correct dimension are thus left out). For all data sets, at least one of the six models fitted properly, so there was always a solution present for each dimension indicator. The number of solutions that were thus left out in both categories, for SIFASP-ML-C, are presented in Table 6.1 (the number of

solutions left out for SIFASP–ML–R were in all conditions of similar magnitude). Also in Table 6.1, the number of solutions used are presented, separated into the two conditions of NFac. From Table 6.1, it can be seen that fitting a model with too many factors was very difficult for SIFASP–ML–C. It is remarkable that, when four factors were present, in 113 cases the model with 4 factors failed to converge to a solution with admissible values. For these 113 cases, all dimension indicators tested necessarily failed to indicate the correct number of factors.

Table 6.1: *The number of proper, nonconverged, and nonadmissible solutions for models with* 1 *to* 6 *factors, for SIFASP–ML–C*

Number of factors		1	2	3	4	5	6
Proper solutions:	2 fact	458	467	129	27	0	0
	4 fact	426	437	402	367	53	5
Nonconverged sol.	2 fact	22	**11***	280	419	466	480
	4 fact	54	43	71	**105***	369	453
Nonadmissible sol.	2 fact	0	**2***	71	34	14	0
	4 fact	0	0	7	**8***	58	22

** In these cases the solution with the correct dimension was improper*

For SIFASP–ML–C, with nonconverged and nonadmissible solutions left out (Figure 6.1a), the measure ECVI gave significantly higher percentages of correct dimension indications (81.0%) than the measures COMBI (73.8%), RMSEA (71.0%), and SCDT (67.7%). As can be seen from Figure 6.1a, the measure ECVI almost always had the highest percentage of correct dimension indications of the four dimension indicators, when summed over each condition of each independent variable separately.

For SIFASP–ML–R, with nonconverged and nonadmissible solutions left out (Figure 6.1b), the measure ECVI also had the highest overall success rate, although the differences with the dimension indicators RMSEA and COMBI were not significant.

It can be seen from Figures 6.1c and 6.1d, that leaving out those solutions that did not converge and leaving in the converged solutions that contained nonadmissible values, led to considerably lower success

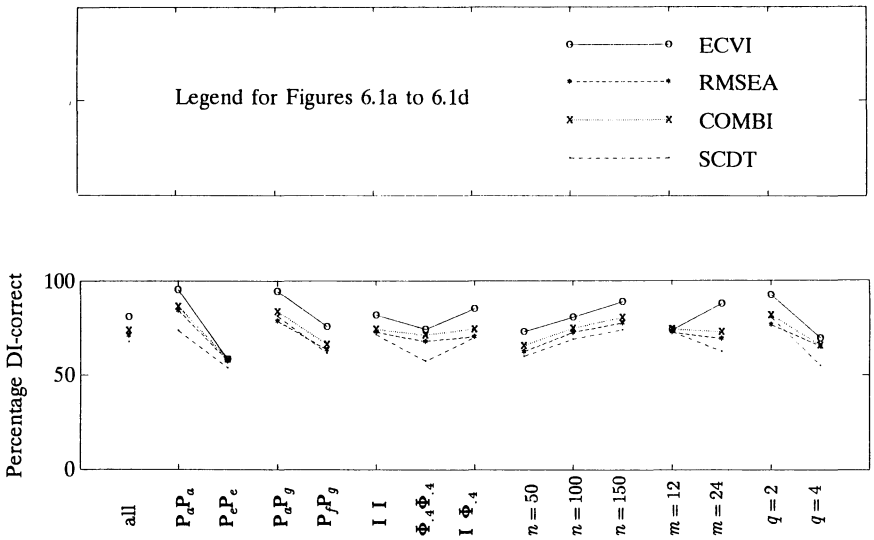


Figure 6.1a *Percentages of correct dimension indications for the method SIFASP – ML – C, averaged over each level of each independent variable, using four different dimension indicators, leaving out all nonconverged and all nonadmissible solutions*

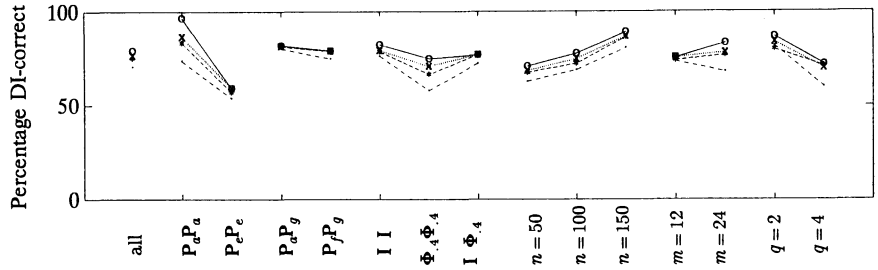


Figure 6.1b *Percentages of correct dimension indications for the method SIFASP – ML – R, averaged over each level of each independent variable, using four different dimension indicators, leaving out all nonconverged and all nonadmissible solutions*

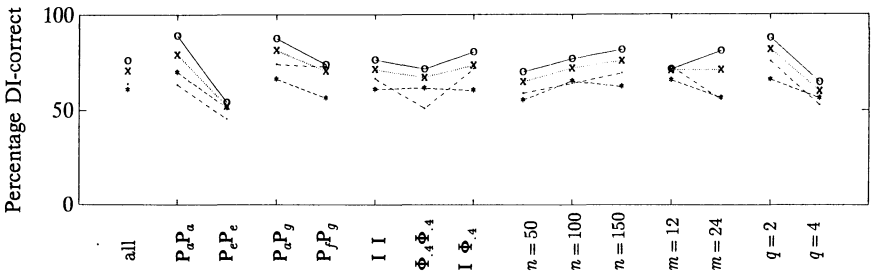


Figure 6.1c *Percentages of correct dimension indications for the method SIFASP–ML–C, averaged over each level of each independent variable, using four different dimension indicators, leaving out all nonconverged solutions*

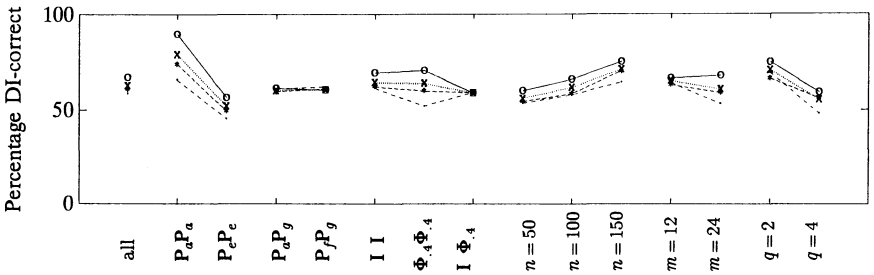


Figure 6.1d *Percentages of correct dimension indications for the method SIFASP–ML–R, averaged over each level of each independent variable, using four different dimension indicators, leaving out all nonconverged solutions*

rates for all dimension indicators. Therefore, leaving out all nonconverged and all nonadmissible solutions is the best course of action.

Answer 1a: The measure ECVI can best be used for dimension indication. For the remainder of this study, only the measure ECVI was used for dimension indication.

Question 1b: "Which of the methods SIFASP–ML and SCA–P is best in indicating the correct dimension?"

Overall, the success rate of the measure ECVI was similar for the two SIFASP–ML methods. SIFASP–ML had a success rate for the measure ECVI of 81.0% when analyzing covariances and of 79.3% when analyzing correlations. The success rate for the measure QDA for SCA–P was 80.1% for the conditions used in this experiment. Because SIFASP–ML had the highest success rate for the measure ECVI when analyzing covariances, the results for this method will be compared with the results for SCA–P. The differences in the success rate of the dimension indicators in samples from one population and samples from two populations was reason to discuss the results for samples from one population and samples from two populations separately.

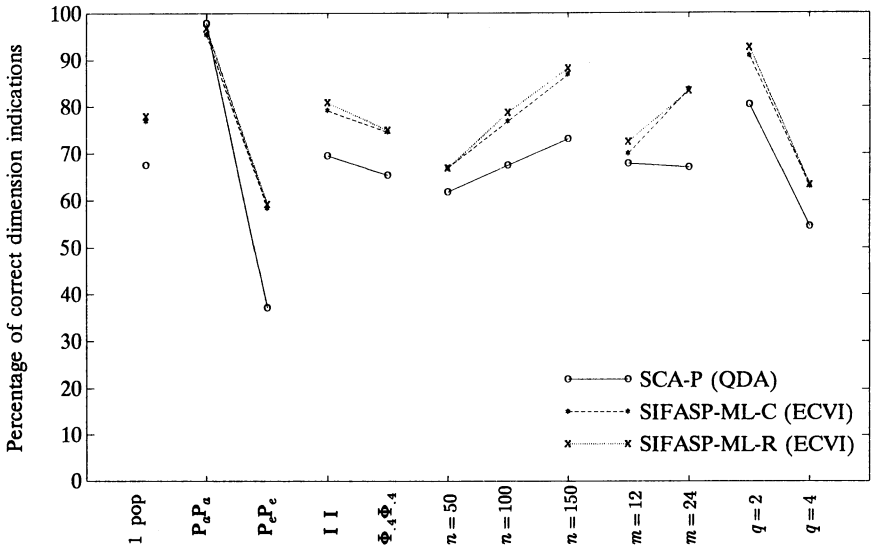


Figure 6.2a *Percentages of correct dimension indications for the two SIFASP – ML – methods (using the dimension indicator ECVI, and leaving out all nonconverged and all nonadmissible solutions) and the method SCA – P (using the dimension indicator QDA), averaged over each level of each independent variable, within the samples from one population*

In Figures 6.2a and 6.2b, the percentages of correct dimension indications by the measure ECVI for SIFASP–ML–C and SIFASP–ML–R, and by the measure QDA for SCA–P are presented, summed over each condition of each independent variable, for samples from one population and samples from two populations, respectively. From Figure 6.2a, it can be seen that, for the samples from *one population*, SIFASP–ML–C and SIFASP–ML–R give similar results for all conditions of the independent variables. The measure ECVI indicates the correct dimension 77.9% of the time for SIFASP–ML–R and 76.9% of the time for SIFASP–ML–C, and the measure QDA indicates the correct dimension 67.5% of the time for SCA–P, although differences are not significant at p<.001: When comparing the results for the measure ECVI for SIFASP–ML–C with the measure QDA for SCA–P by means of a repeated measures ANOVA, there was an effect of Method (p=.005), and interactions of PatternComb × Method (p=.001) and NVar × Method (p=.024), together explaining 38% of the total within subjects variation. The measure ECVI for SIFASP–ML–C has about the same success rates as the measure QDA in condition '$\mathbf{P}_a\mathbf{P}_a$' and higher success rates in condition '$\mathbf{P}_e\mathbf{P}_e$'. The comparison of the results for the measure ECVI for SIFASP–ML–R with the measure QDA for SCA–P led to similar conclusions: There was an effect of Method (p=.003), and an interaction of PatternComb × Method (p=.001), together explaining 33% of the total within subjects variation.

The results for the samples from *two populations* (Figure 6.2b) give a different picture. The measure QDA for SCA–P has a success rate of 92.7%, while the measure ECVI for SIFASP–ML–C has a success rate of 85.2% and the measure ECVI for SIFASP–ML–R has a success rate of 80.6%. We compared the results of SIFASP–ML–C (the best of the SIFASP–methods) to those from SCA–P by a repeated measures anova. The effects of Method, and NVar × Method and the three–way interaction of NFac × NVar × Method were all significant. The NVar × Method interaction pertained to a large difference when there were 12 variables, and a small difference when there were 24 variables. The NFac × NVar × Method interaction can be described as follows: The measure ECVI gave a low success rate of 57.1% in the condition where there were 12 variables and 4 factors and a high

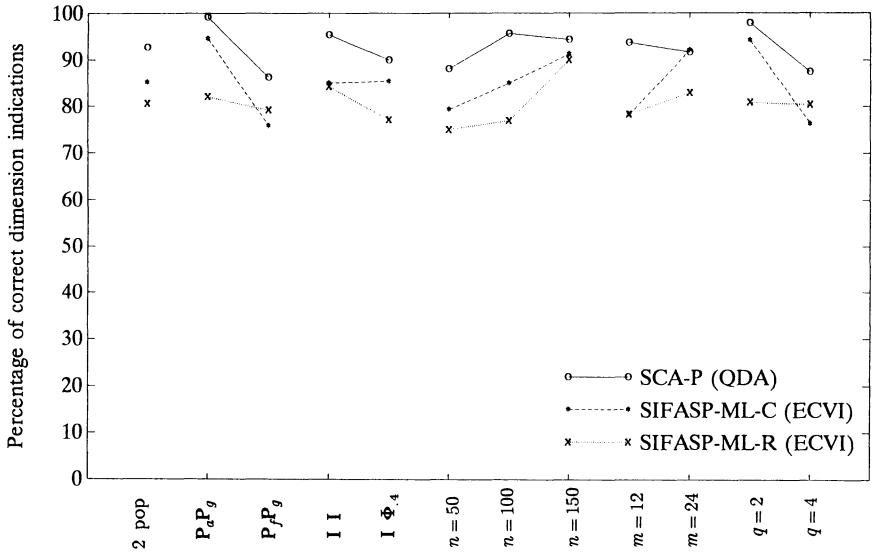


Figure 6.2b *Percentages of correct dimension indications for the two SIFASP-ML-methods (using the dimension indicator ECVI, and leaving out all nonconverged and all nonadmissible solutions) and the method SCA-P (using the dimension indicator QDA), averaged over each level of each independent variable, within the samples from two populations*

success rate of 82.1% in the condition where there were 24 variables and 4 factors, while the measure QDA had success rates of 72.1% and 70.0% in these two conditions, respectively. Thus, this three–way interaction indicates that, an exception to the global tendency of the measure QDA giving higher success rates than the measure ECVI is found in the condition with 24 variables and 4 factors. Finally, there was a significant three–way interaction of PatternComb × NVar × Method. Again it was the measure ECVI that gave a low success rate of 61.7%, this time in the Nvar condition '12 variables' and the PatternComb condition '$\mathbf{P}_f\mathbf{P}_g$', and a high success rate of 90.0% in the NVar condition '24 variables' and the PatternComb condition '$\mathbf{P}_f\mathbf{P}_g$', while the measure QDA had success rates of 88.3% and 84.2% in these two conditions,

respectively. This interaction thus points to another exception to the global superiority of the measure QDA over the measure ECVI. In the PatternComb condition '$\mathbf{P}_a\mathbf{P}_g$', both measures did not differ much across the two NVar conditions. Together the four effects explained 53% of the total within subjects variation.

Answer 1b: From the results, presented above, it was concluded that SIFASP–ML with correlations and covariances, using the measure ECVI, was overall most successful in indicating the correct number of factors in samples from one population, and SCA–P, using the measure QDA, was most successful in indicating the correct number of factors in samples from two populations, although important exceptions exist.

6.6.2 Determining the method with the best recovery of factors

Question 2a: "Is the highest RR found when analyzing covariances (scores standardized over groups) or when analyzing correlations?"
Question 2b: "Which of the methods SIFASP–ML and SCA–P has the highest RR?"
Question 2c: "What, if any, are the most interesting interaction effects between independent variables and method of analysis for the recovery of factors?"

For calculating the RR, all nonconverged and nonadmissible LISREL solutions were left out. The total number of nonconverged and/or nonadmissible solutions (with the correct dimension) was 97 (out of 960) for SIFASP–ML–R and 126 for SIFASP–ML–C (see Table 6.1). Most nonconverged and nonadmissible cases were found for the PatternComb conditions '$\mathbf{P}_e\mathbf{P}_e$' and '$\mathbf{P}_f\mathbf{P}_g$', and within those conditions for the condition when there were 12 variables and 4 factors, and they had a slightly higher tendency of occurring for small sample sizes. The RR of the factors in the nonconverged (as far as they were available) or nonadmissible solutions, found with SIFASP–ML, lay below the average (about 35% for the cases in the condition '$\mathbf{P}_e\mathbf{P}_e$' and about 50%

(SIFASP-ML-R) and 80% (SIFASP-ML-C) for the cases in the condition '$\mathbf{P}_f\mathbf{P}_g$'), indicating that indeed an optimal solution had not (yet) been found. The absence of a proper solution in 10% of the cases for SIFASP-ML-R and in 13% of the cases for SIFASP-ML-C is a serious shortcoming of SIFASP-ML, especially because the identification constraints, used for the methods, were supposed to lead the methods straight towards the correct solution.

When the improper solutions are ignored (leaving 863 proper solutions for SIFASP-ML-R and 834 for SIFASP-ML-C), SIFASP-ML-C had somewhat higher RR's (98.6%) than SIFASP-ML-R (97.8%). Therefore, SIFASP-ML-C is compared with SCA-P. To make the comparison fair, for SCA-P only those cases for which SIFASP-ML-C had a proper solution were selected for the comparison. In these selected cases, the average RR was 99.5%. In the cases left out, the RR of the factors for SCA-P was also below average (RR's for SCA-P were about 77% for the left out cases in the condition '$\mathbf{P}_e\mathbf{P}_e$' and about 95% for the cases in the condition '$\mathbf{P}_f\mathbf{P}_g$'), indicating that in these cases SCA-P also had difficulty in finding the correct solution. Whenever deemed informative, the number of available cases is given. Because in Experiment 1 differences between results for one-population data and two-populations data were found for SCA-P, here also this separation is maintained.

6.6.2.1 *Samples from one population*

In the condition 'samples from one population', there were two conditions of PatternComb. In the PatternComb condition '$\mathbf{P}_a\mathbf{P}_a$', the RR was 100% for all methods (238 cases available). In the PatternComb condition '$\mathbf{P}_e\mathbf{P}_e$' these percentages were 92.6% for SIFASP-ML-R (172 cases), 94.0% for SIFASP-ML-C (167 cases) and 97.9% for SCA-P (167 cases). Because the RR's in PatternComb condition '$\mathbf{P}_a\mathbf{P}_a$' were 100%, they were left out of the analysis.

In Figure 6.3a, the RR's are presented for the selected number of cases for each method, summed over each condition of each independent variable within the PatternComb condition '$\mathbf{P}_e\mathbf{P}_e$'. From Figure 6.3a, we

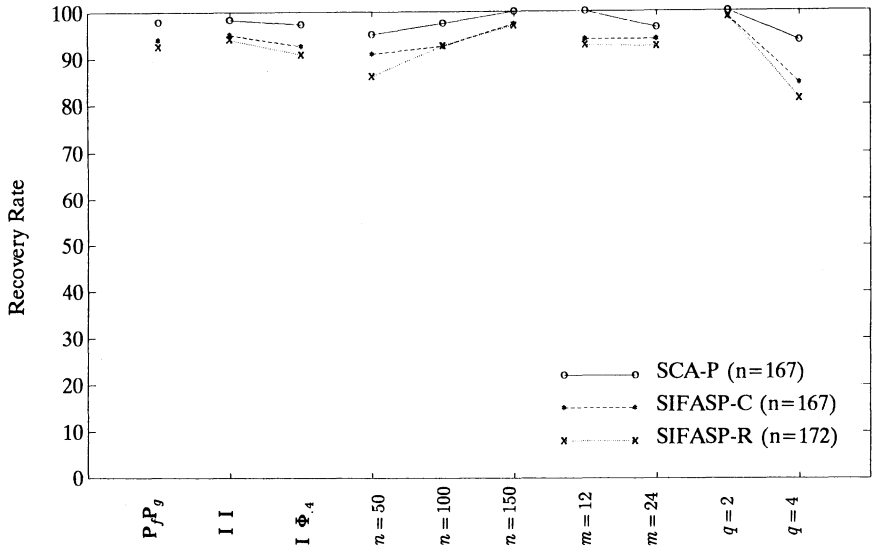


Figure 6.3a *Recovery Rates for the two SIFASP–ML–methods (leaving out all nonconverged and all nonadmissible solutions) and the method SCA–P (for the same selection of data sets as used for the method SIFASP–ML–C), averaged over each level of each independent variable, within PatternComb condition* '$P_e P_e$'

see that there were only small differences between SIFASP–ML–C and SIFASP–ML–R. SCA–P always had a substantially higher RR than SIFASP–ML–C and SIFASP–ML–R, summed over each condition of each independent variable.

Comparing SIFASP–ML–C and SCA–P, it was found that SCA–P had a significantly higher RR than SIFASP–ML–C. (the Method effect explained 19% of the total within subjects variation). Other relevant effects were the interactions of NVar × Method (explaining 7%) and NFac × Method (explaining 13%).

As was done in Experiment 1 with the SCA–methods, in this experiment it was investigated whether SIFASP–ML–C showed preference for either recovery of strong factors or weak factors. For this, the RR's of the strong and the weak factors in PatternComb condition '$P_e P_e$', for the

condition with 24 variables and 4 factors, were calculated for the proper solutions. In total, 42 out of 60 solutions could thus be included, leading to a maximum of (42×4=) 168 strong and 168 weak factors to be recovered. SIFASP–ML–C recovered 100% of the strong factors and only 71% of the weak factors. So for SIFASP–ML–C, recovery of strong factors prevails over recovery of weak factors, as was the case for SCA–P (see Table 5.9, bottom row). Note, however, that recovery of weak factors is higher for SCA–P than for SIFASP–ML–C.

6.6.2.2 *Samples from two populations*

In the PatternComb condition '$\mathbf{P}_a\mathbf{P}_g$', the RR was 100% for all methods (238 cases available). In the PatternComb condition '$\mathbf{P}_f\mathbf{P}_g$' these percentages were 97.2% for SIFASP–ML–R (215 cases), 99.2% for SIFASP–ML–C (191 cases) and 99.7% for SCA–P (191 cases). Because the RR's in PatternComb condition '$\mathbf{P}_a\mathbf{P}_g$' were 100%, they were left out of the analysis.

In Figure 6.3b, the RR's are presented, summed over each condition of each independent variable within the PatternComb condition '$\mathbf{P}_f\mathbf{P}_g$'. All methods have very high RR's, causing a small total within subjects variation. The analysis of variance revealed a significant effect of Method, and significant interactions of NVar × Method, NFac × Method, and NVar × NFac × Method (together explaining about 50% of the total within subjects variation). The three–way interaction of NVar × NFac × Method is explained as follows: All three methods have a RR of 100% or almost 100% when there are 2 factors and 12 or 24 variables present. When there are 4 factors present, SCA–P still has a high RR of 99.6% when there are 24 variables and 98.5% when there are 12 variables, SIFASP–ML–C has a high RR of 99.8% when there are 24 variables and a RR of only 91.9% when there are 12 variables, and SIFASP–ML–R has a RR of 96.7% when there are 24 variables and a RR of only 80.1% when there are 12 variables. So the condition causing the interaction is when there are 12 variables and 4 factors present. In this condition, factors are defined by 3 variables only. It can also be deduced that the SIFASP–methods have trouble in this

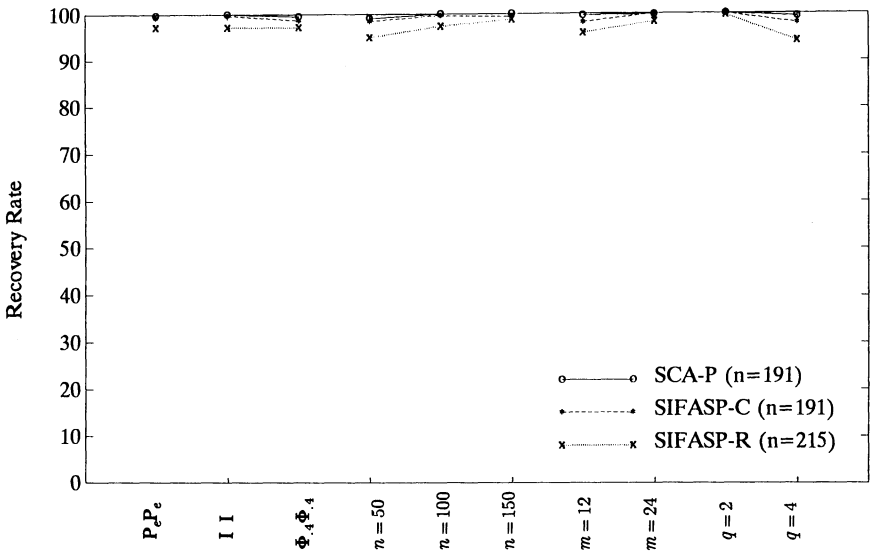


Figure 6.3b *Recovery Rates for the two SIFASP–ML–methods (leaving out all nonconverged and all nonadmissible solutions) and the method SCA–P (for the same selection of data sets as used for the method SIFASP–ML–C), averaged over each level of each independent variable, within PatternComb condition* '$\mathbf{P}_f\mathbf{P}_g$'

condition from the fact that in this condition only 17 proper solutions were available for both SIFASP–methods, against 57 or more in the other three conditions inspected here.

Answer 2a: The question whether the highest RR is found when analyzing covariances or when analyzing correlations can be answered both ways. Analyzing covariances gives higher RR's than analyzing correlations. On the other hand, analysis of covariances gave fewer proper solutions than analysis of correlations. We prefer the analysis of covariances, because in practice one can check if a solution is proper, and for proper solutions the RR from analysis of covariance matrices is higher than that from analysis of correlation matrices; if the solution happens to be

improper it is often discarded anyway.
Answer 2b: SIFASP-ML-C has relatively more trouble in recovering the factors than SCA-P when the number of variables is small (12 instead of 24) and when the number of factors is large (4 instead of 2). It is in these conditions that SIFASP-ML tends to have difficulty in arriving at a proper solution.
Answer 2c: SIFASP-ML-C has a lower RR than SCA-P. Taking also into account that SCA-P had reasonable RR's for the cases for which SIFASP-ML-C did not arrive at a solution, the conclusion must be that SCA-P seems the preferable method.

6.6.3 Determining the method with the best recovery of correlations between factors

Question 3a: "Is the smallest DFC found when analyzing covariances (scores standardized over groups) or when analyzing correlations?"
Question 3b: "Which of the methods SIFASP-ML and SCA-P has the lowest DFC?"
Question 3c: "What, if any, are the most interesting interaction effects between independent variables and method of analysis for the recovery of correlations between factors?"

For calculating the DFC, all nonconverged and nonadmissible LISREL solutions were left out, just as in the case for the success criterion RR. The DFC of the factors in the available nonconverged or nonadmissible solutions, found with SIFASP-ML, lay above the average (.15 in condition '$\mathbf{P}_a\mathbf{P}_a$' for SIFASP-ML-R (average of 2 cases) and SIFASP-ML-C (2 cases); .45 in condition '$\mathbf{P}_e\mathbf{P}_e$' for SIFASP-ML-R (68 cases) and .47 for SIFASP-ML-C (73 cases); .20 in condition '$\mathbf{P}_a\mathbf{P}_g$' for SIFASP-ML-R (2 cases) and .16 for SIFASP-ML-C (2 cases); and .30 in condition '$\mathbf{P}_f\mathbf{P}_g$' for SIFASP-ML-R (25 cases) and .19 for SIFASP-ML-C (49 cases)), indicating that indeed for most of these cases an optimal solution had not (yet) been found.

In Table 6.2, the average DFC values for SIFASP-ML-R, SIFASP-ML-C

and SCA–P are given, for all samples with proper solutions taken together, for each combination of PatternComb and PhiComb, and for each condition of the remaining independent variables. It is also indicated to how many proper (converged and admissible) solutions each value pertains. To make the comparison of the results for the best of the SIFASP–ML methods (SIFASP–ML–C) with results for SCA–P fair, the average DFC values for SCA–P are calculated on the basis of the same (number of) cases as for SIFASP–ML–C. For SCA–P, for the cases thus left out (SCA–P had a proper solution for these cases), the average DFC value in PhiComb condition '**I I**' (56 cases) was of similar magnitude as the average DFC value for SCA–P for the cases for which the SIFASP–methods had a proper solution (see Table 6.2), but the average DFC value in PhiComb conditions '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$' (46 cases) and '$\mathbf{I}\ \mathbf{\Phi}_{.4}$' (24 cases) were about .07 and about .025 higher, respectively, indicating that for the left out cases from these last two PhiComb conditions, SCA–P had more trouble recovering the correct correlations between factors.

From Table 6.2, it can be seen that for all conditions of the independent variables, the differences between SIFASP–ML analyzing covariances and analyzing correlations are negligible, although SIFASP–ML–C has a lower mean DFC value in most of the conditions in Table 6.2. The differences between SIFASP–ML–C and SCA–P are also small, SCA–P almost always having a smaller mean DFC. However, there is no significant effect of Method (p=.158), when comparing these two methods, but there are significant interactions of PatternComb × Method, PhiComb × Method, and a three–way interaction of PatternComb × PhiComb × Method, together explaining 34% of the total within subjects variation. SIFASP–ML–C is somewhat better at recovering correlations of .4 than recovering a correlation of zero between factors. For SCA–P this only holds for the PatternComb condition '$\mathbf{P}_a\mathbf{P}_a$'. For the PatternComb condition '$\mathbf{P}_e\mathbf{P}_e$', the opposite is the case, although when looking at the total number of cases, rather than at the selection of cases for which SIFASP–ML–C gave proper solutions, for SCA–P in the two PhiComb conditions within the PatternComb condition '$\mathbf{P}_e\mathbf{P}_e$', this difference in mean DFC values is not present anymore.

Table 6.2: *Mean DFC values for all methods, for all proper solutions, and within each condition of each independent variable (number of proper cases parenthesized)*

	SIFASP-ML-R	SIFASP-ML-C	SCA-P
All cases	.128 (863)	.122 (834)	.097 (834)
$\mathbf{P}_a\mathbf{P}_a$ and **I I**	.103 (119)	.103 (119)	.092 (119)
$\mathbf{P}_a\mathbf{P}_a$ and $\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	.083 (119)	.083 (119)	.077 (119)
$\mathbf{P}_e\mathbf{P}_e$ and **I I**	.201 (93)	.212 (92)	.084 (92)
$\mathbf{P}_e\mathbf{P}_e$ and $\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	.179 (79)	.163 (75)	.146 (75)
$\mathbf{P}_a\mathbf{P}_g$ and **I I**	.114 (120)	.106 (119)	.095 (119)
$\mathbf{P}_a\mathbf{P}_g$ and **I** $\mathbf{\Phi}_{.4}$	.100 (118)	.097 (119)	.105 (119)
$\mathbf{P}_f\mathbf{P}_g$ and **I I**	.128 (109)	.127 (94)	.095 (94)
$\mathbf{P}_f\mathbf{P}_g$ and **I** $\mathbf{\Phi}_{.4}$	.151 (106)	.121 (97)	.098 (97)
Sampsize = 50	.161 (263)	.154 (253)	.116 (253)
Sampsize = 100	.123 (296)	.119 (285)	.097 (285)
Sampsize = 150	.105 (304)	.097 (296)	.081 (296)
NVar = 12	.136 (404)	.125 (379)	.101 (379)
NVar = 24	.121 (459)	.120 (455)	.094 (455)
NFac = 2	.127 (470)	.126 (467)	.100 (467)
NFac = 4	.129 (393)	.118 (367)	.093 (367)

Answer 3a: About equal DFC values are found when analyzing covariances (scores standardized over groups) and correlations.

Answer 3b: SIFASP-ML overall gives somewhat higher DFC values than SCA-P, but the differences are small. SCA-P also gives good or reasonable DFC values (compared to the cases included in the present comparison) in the cases for which SIFASP-ML has no proper solution.

Answer 3c: SIFASP-ML is better at recovering correlations of .4 between factors, than recovering zero correlations. For SCA-P, there is not a clear effect of underlying correlations on recovery.

6.6.4 Frequency of bad fit for the solution with the correct dimension

Question 4: "How often do the solutions for the correct dimension have bad fit?"

As a check on poor fit, it was inspected how many times the solution with the correct dimension had bad fit, according to the p–value of the chi–square values of the fitted model and the p–value of the RMSEA measure (see Sections 4.3.5 and 4.3.3, formula (4.14)). If one of these p–values would indicate that a model has bad fit, while it is actually the correct model, it can be concluded that the measure fails to recognize the correct model. The percentages of failure are given in Table 6.3.

Table 6.3: *Percentages of times the correct model had bad fit, according to the chi–square value and the measure RMSEA*

	SIFASP–ML–R		SIFASP–ML–C	
	2 fact	4 fact	2 fact	4 fact
Chi–Square				
One population	0.0%	0.0%	0.0%	0.0%
Two populations	62.5%	42.1%	0.0%	0.0%
P(RMSEA<.05)				
One population	0.0%	0.0%	2.9%	0.0%
Two populations	0.0%	0.0%	0.0%	0.0%

From Table 6.3, it can be seen that the correct model usually has reasonable fit, except according to the chi–square value for SIFASP–ML–R in the condition 'samples from two populations'. The overall percentage of times the correct model does not have reasonable fit for 'samples from two populations' is about 50%. Apparently, the chi–square value can not be trusted in this case. These results support Jöreskog's remark (1971, p. 418) that it is not permissible to use χ^2–tests when analyzing

correlation matrices, because in that case the assumptions used in χ^2–testing are violated.

6.7 CONCLUSION

In the present experiment, it was found that SIFASP–ML performs better when covariances are analyzed than when correlations are analyzed. The measure ECVI can best be used for dimension indication for SIFASP solutions, because this measure had the highest percentage of correct dimension indication of the fit indices tested. The measure ECVI, using SIFASP–ML with covariances, was less successful in indicating the correct dimension than SCA–P for samples from two populations, but more successful for samples from one population (specifically in PatternComb condition '$\mathbf{P}_e\mathbf{P}_e$').

In situations where the factors are strongly defined (condition '$\mathbf{P}_a\mathbf{P}_a$') or where differences between the two populations from which samples are drawn are small (condition '$\mathbf{P}_a\mathbf{P}_g$'), recovery of factors is almost 100%. (Note that this differs from the reported exact 100%, because that was based on the 238 (out of 240) proper solutions only). In situations where the factors are weakly defined (condition '$\mathbf{P}_e\mathbf{P}_e$') or where the factor loadings are different across groups (condition '$\mathbf{P}_f\mathbf{P}_g$'), SIFASP–ML–C failed to arrive at a proper solution in about 25% of the cases (for SIFASP–ML–R this is about 20%). When looking at the cases for which a proper solution was found, SIFASP–ML–C had a slightly higher recovery rate than SIFASP–ML–R, although, admittedly, overall recovery of the correct factors was high for both methods, in this situation. SCA–P, however, had a higher recovery rate than SIFASP–ML–C. Because of this higher recovery rate and because SCA–P always arrives at a solution, it seems the preferable method. Looking a bit closer at the differences (in conditions '$\mathbf{P}_e\mathbf{P}_e$' and '$\mathbf{P}_f\mathbf{P}_g$') between SCA–P and SIFASP–ML–C, it was found that the latter method had more trouble in recovering the factors than SCA–P, especially when the number of variables is small (12 instead of 24) and when the number of factors is large (4 instead of 2). It is also in these conditions that SIFASP–ML–C tends to have difficulty in arriving

at a proper solution.

About equal DFC values were found when analyzing covariances and correlations. SIFASP–ML is better at recovering correlations of .4 between factors, than recovering zero correlations. For SCA–P, there is not a clear effect of underlying correlations on recovery. Overall, SIFASP–ML gives somewhat higher DFC values than SCA–P, but differences between the methods are small.

CHAPTER 7

EXPERIMENT 3, COMPARISON OF THE SCA-METHODS, WHEN THE STRENGTH OF FACTORS WITH THE SAME INTERPRETATION DIFFERS ACROSS GROUPS

7.1 INTRODUCTION

In the third experiment, conducted in this study, the objective was to differentiate between the three SCA-methods SCA-W, SCA-P and SCA-S, when the factors differ in strength across groups, while maintaining the same interpretation. That is, in each group, the same variables pertain to a factor, and within each group all variables pertaining to a factor have the same true loading, but the sizes of the loadings for the variables pertaining to a certain factor differ across groups.

For the comparison, three conditions of the independent variable PatternComb were chosen for two groups (Experiment 3a), and one condition of PatternComb was chosen for three groups (Experiment 3b), to simulate, for instance, a longitudinal development, where some factors grow stronger and some factors grow weaker over time. In the analysis of the results, a new success criterion (described in Section 7.3.2) was used in addition to those used in the previous experiments.

7.2 MANIPULATIONS OF THE DATA

The following values were chosen for the independent variables. In the simulated data sets in Experiment 3a there were two samples (p=2), in Experiment 3b there were three samples (p=3). In both experiments there were 12 or 24 variables ($=m$), 2 or 4 factors ($=q$) and sample sizes of n=50, n=100 or n=150. The true pattern matrices ($\mathbf{P}_b$, $\mathbf{P}_d$, $\mathbf{P}_e$, $\mathbf{P}_m$, $\mathbf{P}_n$, $\mathbf{P}_p$ and $\mathbf{P}_q$) used in Experiments 3a and 3b are given in Appendix A. There were three conditions of PatternComb in Experiment 3a and one in Experiment 3b, and, in both experiments there were two different conditions of

PhiComb ($\mathbf{\Phi}_t=\mathbf{I}$ and $\mathbf{\Phi}_t=\mathbf{\Phi}_{.4}$; see Section 4.1.2). Hence, in total there were (3×2×2×2×3=) 72 different conditions in Experiment 3a and (1×2×2×2×3=) 24 different conditions in Experiment 3b to be analyzed. In each condition, ten replications of simulated data were used. The methods of analysis used were SCA–W, SCA–P, SCA–S.

The conditions of PatternComb were chosen such that the strength of the factors differs across groups. In the conditions using pattern combination '$\mathbf{P}_e\mathbf{P}_m$' there are both strong and weak factors present in each group. The strong factors in the first group are weak in the second, and the weak factors in the first group are strong in the second group. In condition '$\mathbf{P}_d\mathbf{P}_n$' the same goes for strong factors and intermediate factors, and in condition '$\mathbf{P}_p\mathbf{P}_q$' the same goes for intermediate factors and weak factors.

In Experiment 3b, there is one PatternComb condition: '$\mathbf{P}_e\mathbf{P}_b\mathbf{P}_m$'. The patterns reflect the longitudinal process simulated by this condition of PatternComb: In the first group ('$\mathbf{P}_e$'), the first half of the factors is strong and the second half of the factors is weak (see Appendix A); In the second group ('$\mathbf{P}_b$'), all factors are of intermediate strength; Finally, in the third group ('$\mathbf{P}_m$'), the first half of the factors is weak and the second half of the factors is strong. Looking at these patterns from a longitudinal (developmental) point of view, the first half of the factors has grown less important (less strong) over time, while the second half of the factors has gained importance.

SIFASP–ML was not included in this experiment. This was done, because (in factor analysis) the strength of factors is usually determined by looking at the size of the estimated loadings. For SIFASP–ML (as defined by Jöreskog, 1971), however, the loadings are kept equal across groups, making this approach pointless. Using the percentage of explained variance of each factor (as is done in component analysis) was considered an approach too far removed from what is usual in SIFASP. For these reasons, it was decided that SIFASP–ML was not a suitable method for the conditions used in the present experiment.

7.3 DEPENDENT VARIABLES

For each data set of two groups, the amount (percentage) of variance explained was calculated for 1, 2, 3, 4, 5 and 6 components drawn. As in Experiments 1 and 2, in the 4 factor condition, the explained variance for 7 components drawn was also calculated.

Because of the results of Experiment 1, only the measure QDA was used for dimension indication. For each solution with the correct dimension, the measures RR and DFC were calculated (see Section 4.4.1. and Section 4.4.2).

7.3.1 A problem with the method SCA-S in some of the conditions studied

In the present experiment, where the effect of different strengths of factors within and across groups is manipulated, it is crucial that the true factors have the same interpretation across groups, because otherwise factor strengths are not comparable. In the PatternComb conditions, used in the present experiment, the nonzero loadings differed both within and across groups, causing that for all samples in the condition where factors are correlated (PhiComb condition '$\Phi_{.4}\Phi_{.4}$', used in half of the data sets studied), the true pattern matrices had perfectly congruent columns, while the columns of the true structure matrices were far from perfectly congruent. (In contrast, when PhiComb condition '**I I**' is used, the columns of the true structure matrices are perfectly congruent as well). This condition is similar to the conditions studied for two-or-more-populations data in Experiment 1. In that experiment, when factors in two-or-more-populations data were correlated, the columns of the true pattern matrices also had higher congruence values than the columns of the true structure matrices. However, in the present experiment, factors differ in strengths within and across groups as well, which causes a serious problem for SCA-S in PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$' in Experiment 3a and, to a lesser degree in PatternComb condition '$\mathbf{P}_e\mathbf{P}_b\mathbf{P}_m$' in Experiment 3b. In PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$', the

factors in one group have, in terms of the true structure matrices, a congruence value with the factors in the other group, with which they should share the same interpretation, of .82 when there are 2 factors and of .77 when there are four factors present. However, the congruence value of the factors in one group with the *other* factors in the other group is .99 (!) when there are 2 factors and .89 when there are four factors present. In this condition we expect SCA–S to encounter serious difficulties in recovering the correct factors, in terms of the true structure matrices. In the other two PatternComb conditions in Experiment 3a, the factors, in terms of the true structure matrices, in one group do have the highest congruence value with the factors in the other group, with which they should share the same interpretation, so there should be no problem for SCA–S in those conditions. Therefore, SCA–S, in which the structure matrices are forced to be columnwise proportional, is not a proper method for describing data in the combination of PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$' and PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$'. The same problem exists in PatternComb condition '$\mathbf{P}_e\mathbf{P}_b\mathbf{P}_m$', used in Experiment 3b, where the same two true pattern matrices are present. However, in PatternComb condition '$\mathbf{P}_e\mathbf{P}_b\mathbf{P}_m$', a third true pattern is present, because there are three groups in that experiment. In PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$' within this PatternComb condition, the factors, based on the true structure matrices for true pattern matrix '$\mathbf{P}_b$', have the highest congruence value with the factors in the two other groups (with true pattern matrices '$\mathbf{P}_e$' and '$\mathbf{P}_m$', respectively), with which they should share the same interpretation, so it is expected that SCA–S will give better results in this condition than it will in the combination of PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$' and PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$' in Experiment 3a, but still will have trouble in retrieving the correct true structure for all three groups.

In practice, however, the true data structure is not known, and there is no a priori reason not to apply SCA–S. Therefore, in the present experiments (3a and 3b), we do apply SCA–S and inspect the recovery of the structure matrices. For comparative purposes, we also inspect the recovery of the pattern matrices for SCA–S in the problematic conditions mentioned above, because, for SCA–S, the pattern matrices are allowed to

differ across groups.

7.3.2 A measure used for inspecting whether the relative strength of the recovered factors resembles the relative strength of the true factors: the Quotient of Component Strengths (QCS)

Besides the question whether the components, found with the SCA–methods, receive the same interpretation as the underlying factors, the question can be asked whether or not the relative strength of the components found is the same as the relative strength of the true factors. To answer this question, a new measure was devised as follows.

The true strength of a factor in the simulated population can be directly derived from the true structure matrix $\mathbf{S}_t$ by taking the sums of squares of the column of this matrix pertaining to that factor. From the description of the loadings in the patterns, used in this experiment (see Appendix A), it can be seen that in each PatternComb condition, there were always factors with two different (true) strengths. We take, for example, the condition where there are two factors present in each group, but of course the same applies in the condition where there are four factors in each group. The quotient, calculated as the strength of the weaker factor divided by the strength of the stronger factor, gives the strength ratio of the true factors in the population.

In the present experiments, for each sample, the strengths of the components, found with each method, were calculated from the structure matrices, found with each method. Next, the strengths of the components matching the weaker true factors are averaged and divided by the average strength of the components, matching the stronger true factors. The strength ratio found with each of the SCA–methods is compared to the strength ratio in the population. Because it is the strength ratio of the *components* found with the SCA–methods that is of interest, this measure is called the Quotient of Component Strengths (QCS). To avoid confusion, the strength ratio of the factors in the population is referred to as the "true QCS" (instead of quotient of factor strengths).

Firstly, it was checked whether or not the QCS, found with each

method, was above or below the true QCS. Secondly, the absolute differences of the QCS and the true QCS were compared. Thirdly, as a combination of the RR measure and the QCS measure, the absolute differences of the QCS and the true QCS were compared when only those cases were included in which all factors were retrieved. That is, all data sets for which the RR is not 100%, were left out of the analysis, because it seems a useless course of action to inspect whether components that do not receive the correct interpretation do have the same relative strengths as the correct factors.

7.4 ANALYSIS

All the measures presented were analyzed with repeated measures analyses of variance. Sums of squares of main and interaction effects were calculated to give an impression of the size of the effects, and the associated averaged univariate F tests were employed for significance testing. In each of the separate analyses, only first and second order interactions among between subjects factors were included in the analysis. All precautions and adjustments, as presented in Section 5.4, were applied.

For the analysis of the results for the dimension indicator QDA, PCA–sep was included as a condition of the within subjects factor, so in that case the within subjects factor had 4 conditions. Outcomes concerning correct dimension indication were aggregated over the ten replications within each condition to obtain a reasonably continuous variable, in the same way as in Experiment 1.

7.5 QUESTIONS TO BE ANSWERED BY THIS EXPERIMENT

The present experiment was conducted to get answers as to what circumstances play a role in the success and failure of the SCA–methods in the recovery of factors and their relative strengths, in conditions

where factors have considerably different strengths across and within groups. Answers will be given in separate subsections of Section 7.6 (Experiment 3a) and Section 7.7 (Experiment 3b), where the question numbers refer to the subsections in which answers to these questions will be given. Summarizing conclusions from the results of this experiment will be drawn in Section 7.8.

1a.) Which of the SCA–methods has the highest success rate on the dimension indicator QDA when factors differ in strength over samples?
1b.) Are there interesting interaction effects between independent variables and method of analysis for the success rate of the dimension indicator QDA?
2a.) Which of the SCA–methods can be judged as best capable of recovering the correct interpretation of factors when factors differ in strength over samples?
2b.) Are there interesting interaction effects between independent variables and method of analysis for the recovery of factors?
3a.) Which of the SCA–methods can be judged as best capable of recovering the relative strength of the factors when factors differ in strength over samples?
3b.) Are there interesting interaction effects between independent variables and method of analysis for the recovery of the relative strength of factors?
4a.) Which of the SCA–methods can be judged as best capable of recovering the correlations between factors when factors differ in strength over samples?
4b.) Are there interesting interaction effects between independent variables and method of analysis for the recovery of correlations between factors?

7.6 RESULTS EXPERIMENT 3A

In the present experiment, a significance level α=.001 was used.

Whenever differences between dependent variables or methods are reported, or when interaction effects are said to occur, effects were significant at α=.001. In the following sections, each of the four main questions, asked at the end of Section 7.5, will be answered separately. At the start of each section, the questions will be repeated.

7.6.1 Finding the SCA-method with the best dimension indication

Question 1a: Which of the SCA-methods has the highest success rate on the dimension indicator QDA when factors differ in strength over samples?
Question 1b: Are there interesting interaction effects between independent variables and method of analysis for the success rate of the dimension indicator QDA?

As mentioned above, the dimension indicator QDA, found to be the dimension indicator with the highest percentage of correct dimension indication in Experiment 1, was used here. A brief comparison of the present results with all dimension indicators used in Experiment 1 (KA1, PA, QA and QDA), led to the same conclusions as in Experiment 1, thus justifying the decision to use only the dimension indicator QDA.

In Figure 7.1, the percentages of correct dimension indication by the dimension indicator QDA for the three SCA-methods are presented. Overall, SCA-P had the highest percentage of correct dimension indication (84.0%), followed by SCA-S (82.4%) and SCA-W (70.7%). The difference between SCA-P and SCA-S was not significant (p=.149), but the difference between either of these methods and SCA-W was. From this it may be clear that there was a large effect of Method. In fact, it explained 20% of the total within subjects variation. In contrast, in Experiment 1, SCA-P also had the highest percentage of correct dimension indication, but SCA-W followed very closely and SCA-S fell behind.

There were substantial and significant interactions of PatternComb × Method, PhiComb × Method, and NFac × Method, together explaining a further 34% of the total within subjects variation. The form of these interactions can be read from Figure 7.1. Specifically, it was found that

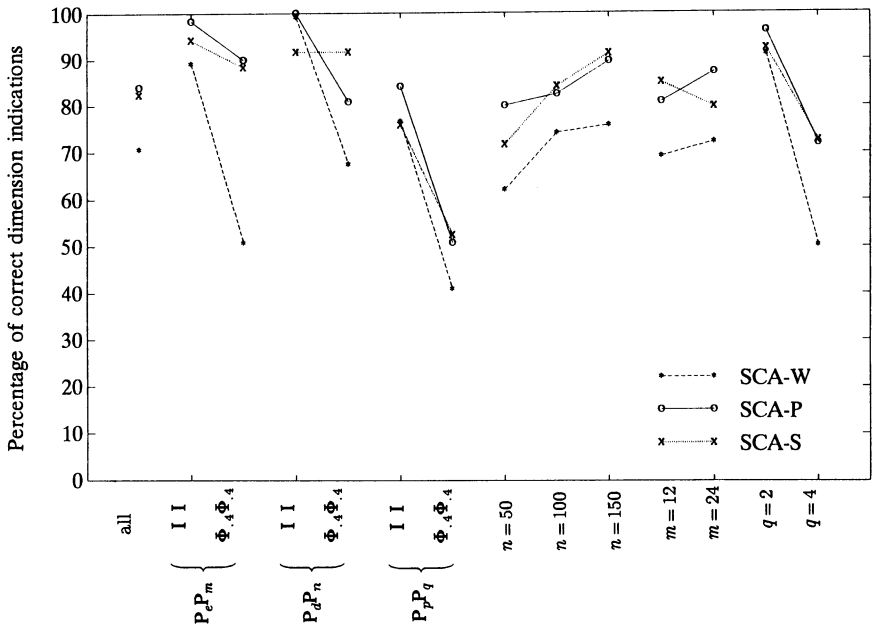


Figure 7.1 *Percentages of correct dimension indications for the three SCA–methods, averaged over (combinations of) the levels of the independent variables, using the dimension indicator QDA*

SCA–W had a relatively low percentage of correct dimension indication in PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$', where factors are strong and weak in the two groups; in PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$', where factors are strongly oblique; and when there are four factors present instead of two. The most interesting difference with the results in Experiment 1 is that here it is SCA–W that mainly causes the interactions, whereas in Experiment 1, SCA–S caused interactions.

Answer 1a: The methods SCA–P and SCA–S have comparable success rates on the dimension indicator QDA when factors differ in strength over samples. SCA–W clearly falls behind.

Answer 1b: SCA–W especially falls behind the other methods when there are

both strong and weak factors present (where the strong and weak factors trade place over groups), when factors are strongly oblique, and when there are four factors present instead of two.

7.6.2 Finding the SCA-method with the best recovery of factors

Question 2a: Which of the SCA-methods can be judged as best capable of recovering the correct interpretation of factors when factors differ in strength over samples?
Question 2b: Are there interesting interaction effects between independent variables and method of analysis for the recovery of factors?

Overall, recovery is high for all methods, and comparable to results found for similar conditions in Experiment 1. In Figure 7.2, the RR's are presented for all the conditions of the independent variables. For SCA-S the RR, based on both the pattern matrices and the structure matrix, are presented, for reasons explained in Section 7.3.1. First we will look at these two RR's for SCA-S. Overall, the RR for SCA-S based on the pattern matrices is higher (95.6%) than the RR based on the structure matrices (89.8%). However, as can be seen from Figure 7.2, only in the combination of PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$' and PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$', the RR, based on the structure matrix, is lower than the RR, based on the pattern matrix. This was the very condition where we expected SCA-S to perform poorly in terms of recovery of the structure matrix. In all other conditions the structure matrix is to be preferred, and will therefore be employed in the sequel. We will now compare the three SCA-methods. SCA-P has the highest RR (97.3%), followed by SCA-W (93.0%) and SCA-S, using the structure matrix (89.8%). The differences between the three SCA-methods were all significant. There was a significant effect of Method, and there were significant and substantial interactions of PatternComb × Method, PhiComb × Method, and a significant and substantial three-way interaction of PatternComb × PhiComb × Method, together explaining 64% of the total within subjects variation. The deviant behavior of the RR for SCA-S in PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$'/PhiComb

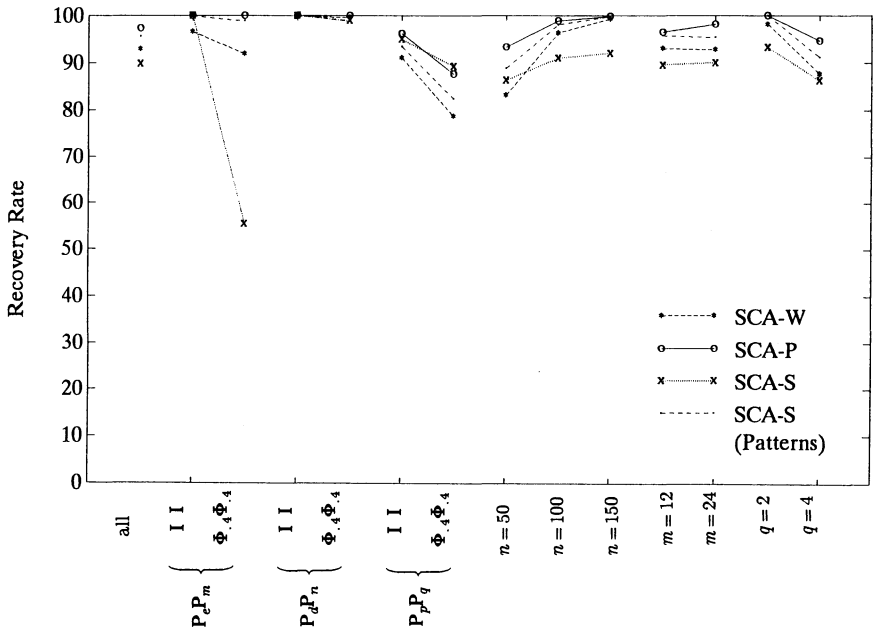


Figure 7.2 *Recovery Rates for the three SCA–methods (and a second time for the method SCA–S, when based on the pattern matrix), averaged over (combinations of) the levels of the independent variables*

condition '$\Phi_{.4}\Phi_{.4}$' clearly accounts for the largest of the interactions. SCA–W especially has lower RR's than SCA–P in PatternComb conditions '$\mathbf{P}_e\mathbf{P}_m$' and '$\mathbf{P}_p\mathbf{P}_q$' (the conditions with weak factors).

Comparing the results for the three conditions of PatternComb, with those in similar PatternComb conditions in Experiment 1, we find that RR's are higher for all methods in PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$' than in condition '$\mathbf{P}_e\mathbf{P}_e$', especially for SCA–P and SCA–S, with the exception of PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$'/PhiComb condition '$\Phi_{.4}\Phi_{.4}$' for SCA–S. RR's in PatternComb condition '$\mathbf{P}_d\mathbf{P}_n$' closely match those in condition '$\mathbf{P}_d\mathbf{P}_d$' (RR's close or equal to 100%). The RR's in PatternComb condition '$\mathbf{P}_p\mathbf{P}_q$', finally, take a position in between those in PatternComb conditions '$\mathbf{P}_b\mathbf{P}_b$' and '$\mathbf{P}_c\mathbf{P}_c$', as was expected, considering that the average loadings

in PatternComb condition '$\mathbf{P}_p\mathbf{P}_q$' are also in between those in PatternComb conditions '$\mathbf{P}_b\mathbf{P}_b$' and '$\mathbf{P}_c\mathbf{P}_c$'.

Answer 2a: The methods SCA–P and SCA–S are best capable of recovering factors when factors differ in strength across samples, while maintaining the same interpretation, with the exception for SCA–S in the situation where there are strong and weak factors in both samples that change place over samples and are strongly oblique.
Answer 2b: SCA–W especially falls behind the other methods when there are weak factors present.

7.6.3 Finding the SCA-method with the best recovery of the relative strength of factors

Question 3a: Which of the SCA–methods can be judged as best capable of recovering the relative strength of the factors when factors differ in strength over samples?
Question 3b: Are there interesting interaction effects between independent variables and method of analysis for the recovery of the relative strength of factors?

In Table 7.1, the mean over and underestimation of the true QCS (see Section 7.3.2) is given for each method, together with the number of times over and underestimation occurred.

Table 7.1: *Over and underestimation and mean absolute difference from the true QCS by the SCA–methods*

	SCA–W(cases)	SCA–P(cases)	SCA–S(cases)
Mean overestimation	.16 (674)	.16 (676)	.13 (398)
Mean underestimation	.03 (46)	.03 (44)	.12 (322)
Mean abs. diff.	.15 (720)	.15 (720)	.13 (720)

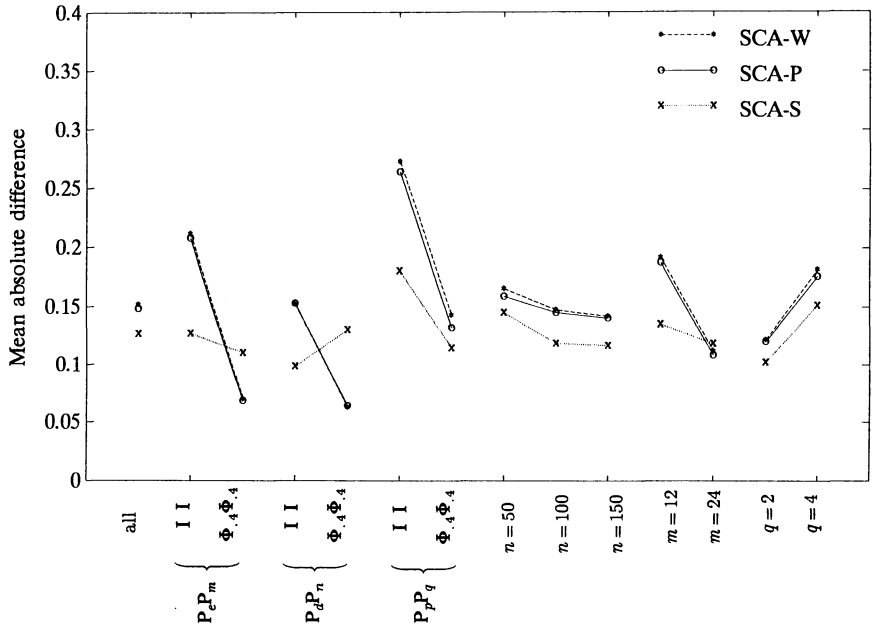


Figure 7.3 *Mean absolute differences of the QCS for each SCA–method and the true QCS, averaged over (combinations of) the levels of the independent variables*

It can be seen from Table 7.1 that SCA–W and SCA–P generally overestimate the QCS. That is, these methods tend to overestimate the relative strength of the weaker components found. SCA–S seems to display no such bias.

In Figure 7.3, the mean absolute differences between the QCS, found with each of the SCA–methods, and the true QCS are presented, when averaged over each condition of each independent variable. There was a significant effect of Method, explaining only 4% of the total within subjects variation, the QCS for SCA–S being the closest to the true QCS. There were substantial and significant interactions of PatternComb × Method and NVar × Method (see Figure 7.3), together explaining 11% of the total within subjects variation. The largest effects were the interaction

of PhiComb × Method and the three–way interaction of PhiComb × NVar × Method, both explaining 21% of the total within subjects variation. The interaction of PhiComb × Method comes from the fact that while the QCS for SCA–W and SCA–P is larger when factors are uncorrelated (.21 for both methods) than when they are correlated (.09 for both methods), for SCA–S there is little difference on the average QCS for uncorrelated and correlated factors (.14 and .12, respectively). The three–way interaction of PhiComb × NVar × Method is described as follows. In PhiComb condition '**I I**', the QCS for *all methods* is smaller when there are 24 variables than when there are 12 variables present (.26 for SCA–W and SCA–P, and .20 for SCA–S at 12 variables, against .16 for SCA–W and SCA–P, and .07 for SCA–S at 24 variables). In PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$', the QCS for SCA–W and SCA–P is smaller when there are 24 variables than when there are 12 variables present (.12 at 12 variables against .06 at 24 variables), but the QCS for SCA–S is larger when there are 24 variables than when there are 12 variables present (.07 at 12 variables against .17 at 24 variables).

All cases are included in Table 7.1, but not for all cases all factors were recovered. This might contaminate the results for the QCS measure. Therefore, the mean absolute difference of the value of the QCS measure for each of the SCA–methods and the true QCS value, for those cases for which each of the SCA–methods had a RR of 100%, was inspected. SCA–W recovered all factors 577 times out of 720 (80.1%), SCA–P 681 times (94.5%) and SCA–S 544 times (75.6%, in terms of the structure matrix). In 485 cases, all three SCA–methods recovered all factors. In the combination of PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$' and PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$', SCA–S had a RR of 100% in only 2 cases out of 120, so this condition is almost never present in this selection. For all other conditions, results led to practically the same conclusions as when all cases were included in the analysis. Mean absolute differences for the QCS measure were .16 (SCA–W), .16 (SCA–P) and .12 (SCA–S).

Answer 3a: SCA–S is most successful in correctly recovering the absolute and relative strength of factors when factors differ in strength over

samples, followed by both SCA–P and SCA–W. However, there are strong and interesting interactions between independent variables and method of analysis.

Answer 3b: When factors are uncorrelated, SCA–S is best at recovering absolute and relative strength of factors, but while higher correlations between the factors and the presence of a larger number of variables strongly enhances the recovery of relative factor strength for SCA–P and SCA–W, this has only a small positive ('$\mathbf{P}_p\mathbf{P}_q$') or negative ('$\mathbf{P}_e\mathbf{P}_m$' and '$\mathbf{P}_d\mathbf{P}_n$') effect for SCA–S.

7.6.4 Finding the SCA-method with the best recovery of correlations between factors

Question 4a: Which of the SCA–methods can be judged as best capable of recovering the correlations between factors when factors differ in strength over samples?

Question 4b: Are there interesting interaction effects between independent variables and method of analysis for the recovery of correlations between factors?

The mean DFC values for each condition of each independent variable are presented in Figure 7.4. Compared to Experiment 1 (condition 'samples from one population'), overall mean DFC values are a little higher for SCA–P and SCA–W and a little lower for SCA–S. As in Experiment 1, the strongest interaction was that of PhiComb × Method, explaining 32% of the total within subjects variation. When factors are uncorrelated, SCA–S has the smallest DFC value (.071 versus .084 for SCA–W and .090 for SCA–P). When factors are correlated, SCA–S has the largest DFC value (.21 versus .15 for SCA–W and .14 for SCA–P).

A substantial effect was found for Method, in contrast to what was found in Experiment 1 in the condition 'samples from one population'. SCA–S had a higher DFC value than SCA–P and SCA–W. Other substantial effects, that were also not found in Experiment 1 in the condition 'samples from one population', were the interaction of PatternComb ×

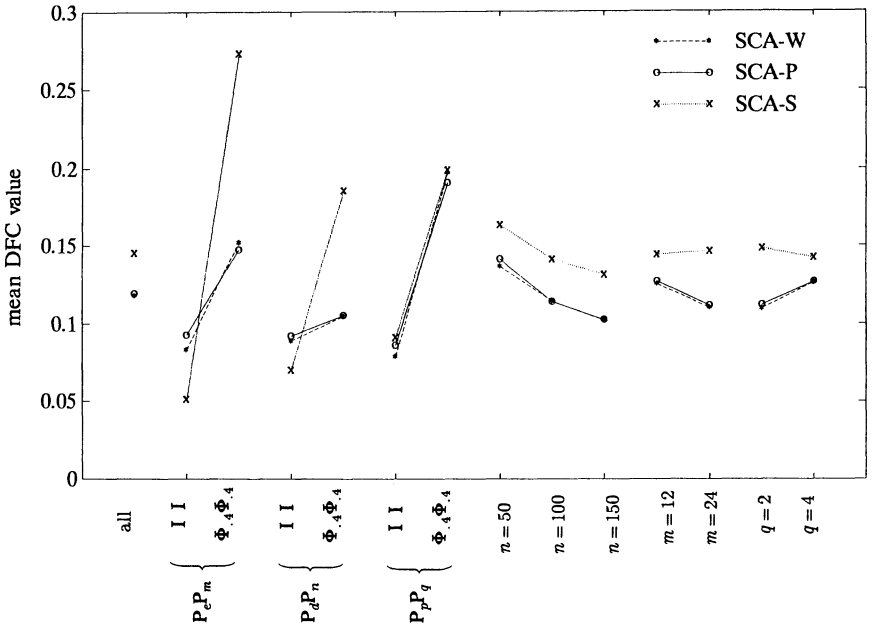


Figure 7.4 *Mean DFC values for the SCA-methods, averaged over (combinations of) the levels of the independent variables*

PhiComb × Method, and the interaction of PhiComb × NFac × Method, together with the Method effect explaining 43% of the total within subjects variation. Within PatternComb condition '$\mathbf{P}_p\mathbf{P}_q$', there was little difference between the mean DFC values of the three SCA-methods in PhiComb conditions '**I I**' and in PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$' (see Figure 7.4), while within the other PatternComb conditions, especially in '$\mathbf{P}_e\mathbf{P}_m$', SCA-S had a smaller mean DFC value than the other SCA-methods in PhiComb condition '**I I**' (.05 vs. .08 for SCA-W and .09 for SCA-P), but higher DFC values than the other SCA-methods in PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$' (.27 vs. .15 for both SCA-W and SCA-P). Within PhiComb condition '**I I**', SCA-W and SCA-P had larger DFC values when there were four factors present (average .09 and .10, respectively) than when there were two factors present (average .08 and .09) and SCA-S had smaller DFC values

when there were two factors present (average .06) than when there were four factors present (average .08). Within PhiComb condition '$\Phi_{.4}\Phi_{.4}$', SCA–S gave high mean DFC values both when there were two factors (.24) and when there were four factors (.20), while SCA–W and SCA–P had smaller mean DFC values when there were two factors present (average .13 for both methods) than when there were four factors present (average .17 for both methods).

Answers 4a and 4b: The SCA–methods are about equally good in recovering the correlations between factors when there are only intermediate and weak factors present, and when the factors are uncorrelated. When there are strong factors present and when the factors are correlated, SCA–P and SCA–W are better able to recover the correlations between factors than SCA–S.

7.7 RESULTS EXPERIMENT 3B

In the present experiment, the same approach towards significance is taken as in previous experiments. In the following sections, the four main questions, asked in Section 7.5, will be answered separately.

7.7.1 Finding the SCA-method with the best dimension indication

SCA–P has the highest percentage of correct dimension indication with the measure QDA (89.6%), followed by SCA–S (85%) and SCA–W (79.2%). The results found lead to the same conclusions as those in Experiment 3a.

7.7.2 Finding the SCA-method with the best recovery of factors

In Table 7.2, the RR's are presented for the three SCA–methods, and a second time for SCA–S, in terms of the pattern matrices.

SCA–P has complete recovery, closely followed by SCA–W. The RR for SCA–S, in terms of the structure matrix, falls behind considerably. As in

Table 7.2: *Recovery rates for the three SCA-methods and for the method SCA-S, when based on the pattern matrices*

	SCA-W	SCA-P	SCA-S structure	SCA-S pattern	n
$\mathbf{P}_e\mathbf{P}_b\mathbf{P}_m$	97.8	100.0	85.9	99.1	240
I I	98.6	100.0	100.0	99.8	120
$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	96.9	100.0	71.9	98.4	120
n=50	93.9	100.0	86.4	97.3	80
n=100	99.7	100.0	85.6	100.0	80
n=150	99.8	100.0	85.8	100.0	80
12 var	98.3	100.0	86.2	99.7	120
24 var	97.3	100.0	85.7	98.5	120
2 fact	99.7	100.0	88.2	99.7	120
4 fact	95.8	100.0	83.7	98.5	120

Experiment 3a, this only happens when the factors are correlated ('$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$'). A closer look at the RR's for each of the three groups ('$\mathbf{P}_e$', '$\mathbf{P}_b$' and '$\mathbf{P}_m$') separately, shows all factors in group '$\mathbf{P}_b$' are recovered, against an average of 54.4% in groups '$\mathbf{P}_e$' and '$\mathbf{P}_m$' (this is about the same percentage as found in Experiment 3a in PatternComb condition '$\mathbf{P}_e\mathbf{P}_m$'/PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$'). The RR's for SCA-S, in terms of the pattern matrices, are very high, taking a position in between SCA-P and SCA-W. These results support the findings in Experiment 3a.

7.7.3 Finding the SCA-method with the best recovery of the relative strength of factors

The major difference of the present experiment with Experiment 3a is that in the present experiment we have three groups instead of two. In each group there were factors of two different strengths, and over the three groups, factors with the same interpretation had three different strengths: weak, intermediate and strong. This meant that there were three measures available for the relative strength. We compared strengths of the weak and the strong factors, the intermediate and the strong

factors, and the weak and the intermediate factors. Mean over- and underestimation, as well as the mean absolute differences for the factors, found with the SCA-methods, for each of the three QCS measures, are given in Table 7.3.

From Table 7.3, it can be seen that SCA-W and SCA-P behave the same for this measure. Both methods overestimated the relative strengths of the weak factors and the intermediate factors, compared to the strong factors in 93% and 95% of the cases, respectively, and overestimate the relative strength of the weak factors compared to the intermediate factors in 70% of the cases. SCA-S underestimates the relative strength of the weak factors about just as often as it overestimates the relative strength of the weak factors. The comparison of the weak and the strong factors (upper section of Table 7.3) leads to the same conclusion as in Experiment 3a.

All cases are included in Table 7.3, but not for all cases all factors were recovered. This might contaminate the results for the QCS measure. Therefore, in Table 7.4, the mean absolute differences for the

Table 7.3: *Mean over- and underestimation, and mean absolute differences from the true QCS for the SCA-methods*

	SCA-W(cases)	SCA-P(cases)	SCA-S(cases)
Weak vs. strong factors			
Mean overestimation	.15 (224)	.14 (224)	.13 (126)
Mean underestimation	.02 (16)	.02 (16)	.11 (114)
Mean abs. diff.	.14 (240)	.14 (240)	.12 (240)
Intermediate vs. strong factors			
Mean overestimation	.13 (229)	.13 (229)	.10 (149)
Mean underestimation	.03 (11)	.03 (11)	.08 (91)
Mean abs. diff.	.13 (240)	.13 (240)	.09 (240)
Weak vs. intermediate factors			
Mean overestimation	.21 (167)	.21 (168)	.21 (120)
Mean underestimation	.12 (73)	.12 (72)	.22 (120)
Mean abs. diff.	.18 (240)	.18 (240)	.21 (240)

Table 7.4: *Mean absolute differences from the true QCS for the SCA-methods, including only the* 109 *cases in which all three SCA-methods had a RR of* 100%

	SCA-W	SCA-P	SCA-S
Weak vs. strong factors			
Mean abs. diff.	.20	.19	.13
Intermediate vs. strong factors			
Mean abs. diff.	.15	.15	.09
Weak vs. intermediate factors			
Mean abs. diff.	.24	.24	.21

components, found with the SCA-methods, for each of the three QCS measures are given, when including only those cases for which all three SCA-methods had a RR of 100%. This was the case in a total of 109 out of 240 cases (45%).

For SCA-W and SCA-P, the mean absolute differences in the cases where all factors are recovered are higher than when averaged over all cases, while for SCA-S there is hardly any difference. So, surprisingly, the results for SCA-W and SCA-P are actually worse when only looking at cases where factors were correctly recovered than when looking at all cases. This was not found in Experiment 3a.

7.7.4 Finding the SCA-method with the best recovery of correlations between factors

The mean DFC values for SCA-W and SCA-P were about equal (.1158 and .1157, respectively). SCA-S had a higher mean DFC value of .1697. Results for the DFC measure led to the same conclusions as in Experiment 3b.

7.8 CONCLUSION

When factors differ in strength across groups, while maintaining the

same interpretation, SCA–P and SCA–S have comparable success rates on the dimension indicator QDA. SCA–W falls behind in some conditions, mentioned in Section 7.6.1.

The methods SCA–P and SCA–S are best capable of recovering factors, except for SCA–S in the special situation where there are strong and weak factors in the (two or more) groups that switch places over groups and are strongly oblique. In the latter case, recovery for SCA–S is exceptionally low. SCA–W performs relatively badly when there are weak factors present.

SCA–S is most successful in correctly recovering the absolute and relative strength of factors, followed by SCA–P and SCA–W, which give about equal results. It must be noted, however, that, in the exceptional condition mentioned above, SCA–S almost never correctly recovers all factors, devaluating the correct recovery of relative factor strength. High correlations between the factors and the presence of a larger number of variables enhances the recovery of relative factor strength for SCA–P and SCA–W, while it has little effect for SCA–S. In the three group condition used in Experiment 3b, SCA–W and SCA–P had a worse recovery of factors strength when averaged over cases with only correctly recovered factors than when averaged over all cases. This result was not found in similar two group conditions in Experiment 3a.

When there are strong factors present and when the factors are correlated, SCA–P and SCA–W are better able than SCA–S to correctly recover the correlations between factors. In the other conditions the SCA–methods perform about equally well.

CHAPTER 8

EXPERIMENT 4, COMPARISON OF THE SCA-METHODS AND SIFASP-ML-C, FOR VARIOUS CONDITIONS

8.1 INTRODUCTION

In the experiments, presented in this chapter, the objective was again to see how SIFASP–ML compared with the three SCA–methods, but now in somewhat more extreme or more complicated conditions. Because it was found in Experiment 2 that SIFASP–ML gave the best results when analyzing scores standardized over groups (from now on denoted as covariances), the method was applied to covariances. The SCA–methods, in turn, are applied to correlations as before. The correlation matrices, analyzed by the SCA–methods, are derived from the same scores as the covariance matrices for SIFASP–ML. In total there are three experiments in this chapter, each one investigating the success of the methods, with conditions that go beyond the scope of Experiment 2.

8.2 MANIPULATIONS OF THE DATA

The manipulations of the data, chosen for the three experiments in this chapter, differ from those in the earlier experiments in that they use a larger number of groups (4a), larger sample sizes (4b), or overlapping factors (4c). These experiments were done to broaden the generalizability of the results from Experiments 1 and 2. In the present section, the specific choices for Experiments 4a to 4c will be presented. The true pattern matrices ($\mathbf{P}_b$, $\mathbf{P}_e$, $\mathbf{P}_g$, $\mathbf{P}_h$, $\mathbf{P}_l$, $\mathbf{P}_r$ and $\mathbf{P}_s$) used in Experiments 4a to 4c are given in Appendix A.

Firstly, for all experiments in this Chapter, factors were chosen to be uncorrelated, to ease comparison. The other choices are discussed experiment by experiment.

For Experiment 4a, where data from four groups were investigated, PatternComb conditions '$\mathbf{P}_b\mathbf{P}_b\mathbf{P}_b\mathbf{P}_b$' and '$\mathbf{P}_b\mathbf{P}_g\mathbf{P}_h\mathbf{P}_l$' were selected, the first condition simulating a situation where all four groups come from the same population and the second condition simulating a situation where all four groups come from different populations. The CM of the true pattern matrix '$\mathbf{P}_b$' and any of the three other patterns is .94, as it was for uncorrelated factors in PatternComb condition '$\mathbf{P}_b\mathbf{P}_g$' in Experiment 1, from which this PatternComb condition can be viewed as an extension. For SIFASP–ML, no comparison with previous results is possible. Furthermore, sample size was fixed at 150, NVar was varied over 12 and 24, and NFac over 2 and 4, leading to a total of eight conditions.

For Experiment 4b, involving larger sample sizes, PatternComb condition '$\mathbf{P}_e\mathbf{P}_e$', with 12 variables and 4 factors, was selected, because there was considerable room for improvement for the SCA–methods and SIFASP–ML on the three success criteria Dimension Indication (DI), RR and DFC, compared to results for this condition in Experiment 2. The sample sizes were taken equal to 300, 500 and 1000.

Finally, for Experiment 4c, two new PatternComb conditions, with overlapping factors, were constructed. It was decided to use 12 variables and 2 factors, and let half the factors overlap, so 4 variables load on two factors. The other four variables were given high ('$\mathbf{P}_r$') or moderate ('$\mathbf{P}_s$') loadings (see Appendix A), making comparison with PatternComb conditions $\mathbf{P}_a$ and $\mathbf{P}_b$, respectively, possible. All nonzero loadings in patterns $\mathbf{P}_r$ and $\mathbf{P}_s$, correspond to a *nsr* of .5 and 1.5, respectively. For the variables have a nonzero loading for two factors, however, this led to a lower true loading, due to rowwise standardization. Because the PatternComb conditions are new, we were interested in differential behavior for different sample sizes and varied it over 50, 100, and 150, thus leading to a total of six conditions.

8.3 DEPENDENT VARIABLES

For the SCA–methods, for each data set (of two or four groups), the

amount (percentage) of variance explained was calculated for 1, 2, 3, 4, 5 and 6 components drawn, and for SIFASP–ML, six models were fitted in which 1, 2, 3, 4, 5 and 6 factors were specified, respectively. The identification constraints employed with these models are given in Appendix B. For each of the six models (if the estimation procedure converged), the values of the fit indices were obtained. As before, for the SCA–methods, in the 4 factor conditions, the explained variance for 7 components drawn was also calculated.

The indicator QDA was used for dimension indication for the SCA–methods and ECVI was used for SIFASP–ML. For each solution with the correct dimension, the measures RR and DFC were calculated (see Section 4.4.1. and Section 4.4.2)

8.4 ANALYSIS

In the present experiments, we do not, contrary to previous experiments, resort to analyses of variance. The main reason for this is that in the present experiments the number of cases was small, and such analyses would have contributed little to the interpretation of the results and to the conclusions.

8.5 QUESTIONS TO BE ANSWERED BY THESE EXPERIMENTS

Experiments 4a and 4b were conducted to get answers as to whether performance of the methods under investigation improves with the addition of groups and subjects, respectively. Experiment 4c was conducted to investigate whether or not factors are better recovered when there is overlap between the factors. Answers will be given in separate subsections of Section 8.6, where the question numbers below refer to the subsections in which answers to these questions will be given. Summarizing conclusions from the results of this experiment will be drawn in Section 8.7. The questions to be answered in this chapter are:

1.) Which of the methods has the highest scores on the success measures Dimension Indication (DI), RR and DFC, when analyzing scores for four groups?
2.) Which of the methods has the highest scores on the success measures DI, RR and DFC, when analyzing scores for large samples?
3.) Which of the methods has the highest scores on the success measures DI, RR and DFC, when factors within each group overlap?

8.6 RESULTS EXPERIMENT 4

8.6.1 Experiment 4a: Four groups

Question 1: Which of the methods has the highest scores on the success measures DI, RR and DFC, when analyzing scores for four groups?

In the present experiment, there was a total of 80 cases, and for the dimension indication the absolute number of correct indications is given instead of percentages. In PatternComb condition '$\mathbf{P}_b\mathbf{P}_b\mathbf{P}_b\mathbf{P}_b$' (40 cases), the measure QDA indicated the correct dimension 40 times for SCA–W and SCA–P, and 39 times for SCA–S. This is similar to the PatternComb condition '$\mathbf{P}_b\mathbf{P}_b$' in Experiment 1 (38, 38 and 39 times, respectively). In PatternComb condition '$\mathbf{P}_b\mathbf{P}_g\mathbf{P}_h\mathbf{P}_l$' (40 cases), the measure QDA indicated the correct dimension 39 times for SCA–W and SCA–P, and 38 times for SCA–S. The measure ECVI indicated the correct dimension 80 times for SIFASP–ML.

For all models with the correct dimension, SIFASP–ML arrived at a proper solution, so no cases were left out for the success criteria RR and DFC. All methods had a RR of 100%. Mean absolute DFC values over the 80 cases were .07, .07, .08 and .07 for SCA–W, SCA–P, SCA–S and SIFASP–ML, respectively. These values differed little (maximum .03) over each of the conditions of the independent variables. All methods gave positive and negative correlations about an equal number of times. In

PatternComb condition '$\mathbf{P}_b\mathbf{P}_b$', in Experiment 2, about the same average DFC values as within PatternComb condition '$\mathbf{P}_b\mathbf{P}_b\mathbf{P}_b\mathbf{P}_b$' were found within each condition of each independent variable (differences were always smaller than .01). Because the differences between the methods on all three success criteria were so small, no significance tests were performed.

Answer 1: For the four groups in the present experiment, the results for the four methods seem to differ very little, and all methods perform very well. It may be noted that the estimates of the correlations between factors do not get better when more groups from the same population are taken.

8.6.2 Experiment 4b: Samples of n=300, n=500 and n=1000

Question 2: Which of the methods has the highest scores on the success measures DI, RR and DFC, when analyzing scores for large samples?

In the present experiment, there was a total of 30 cases, and for the dimension indication the absolute number of correct indications is given instead of percentages. For comparison, results for the same condition ('$\mathbf{P}_e\mathbf{P}_e$', 'I I', 12 variables, 4 factors) from Experiment 2, at sample sizes 50, 100, and 150 are presented as well. The number of correct dimension indications for each method at the six sample sizes are given in Table 8.1. From Table 8.1, it can be seen that, while all methods mostly failed when using small sample sizes (all methods gave underestimations), at sample sizes 300 or more, SIFASP-ML starts indicating the correct number of factors frequently, while the SCA-methods have only moderate success.

For the models with the correct dimension, SIFASP-ML arrived at a proper solution only 40 times out of 60, so 20 cases were left out for the comparison of the success criteria RR and DFC. Especially at small sample sizes, SIFASP-ML failed to arrive at a proper solution more often than it did arrive at a proper solution. However, at a sample size of 500, all solutions were proper. The fact that the SCA-methods generally failed in indicating the correct number of factors can be explained by looking more closely at the true pattern matrix '$\mathbf{P}_e$', used in this

Table 8.1: *Number of correct dimension indications*

Sample Size	SIFASP–ML (ECVI)	SCA–W (QDA)	SCA–P (QDA)	SCA–S (QDA)
50	0	1	2	3
100	0	1	1	1
150	1	0	0	0
300	8	2	2	3
500	9	4	3	3
1000	10	2	2	2

experiment (see Appendix A). In each group there are two strong and two weak factors present. Considering that one can expect the increase in the total amount of variance explained, when increasing the number of components for each method to three and four, to be about equal, this automatically leads to a high QDA value for the 2 factors solution. This is in fact what happened. The measure QDA indicated 2 components 46 times, 46 times, and 45 times for SCA–W, SCA–P and SCA–S, respectively. As a logical progression along this line of thought, one would expect that the QDA measure would give the 4 components solutions a high QDA value as well, so, when looking at the second highest QDA values (so–called second choices of the measure QDA), we indeed found that 4 components were indicated 41 times, 44 times, and 41 times for SCA–W, SCA–P and SCA–S, respectively. The point is that the measure QDA should be used with a little more caution than was done in the present study, because it may contain valuable information about alternative choices for the number of components to retain.

Results for the RR measure for the selection of proper cases are presented in Table 8.2. From Table 8.2, it can be seen that SIFASP–ML has lower RR's than the SCA–methods at all sample sizes except n=1000. For comparison, it should be noted that the RR's for the SCA–methods including all cases were 88%, 88%, and 78% at n=50 for SCA–W, SCA–P and SCA–S, respectively, 89%, 93% and 93% at n=100, respectively, 100%, 100% and 98% at n=150, respectively, and 100% for all SCA–methods at n=300.

Table 8.2: *Recovery rates for the proper solutions*

Sample Size	SIFASP–ML	SCA–W	SCA–P	SCA–S	number of proper solutions
50	67%	92%	100%	75%	3
100	88%	100%	100%	100%	2
150	83%	100%	100%	100%	6
300	89%	100%	100%	100%	9
500	95%	100%	100%	100%	10
1000	100%	100%	100%	100%	10

Results for the DFC measure for the selection of proper cases are presented in Table 8.3. From Table 8.3, it can be seen that for all methods, the true correlations are better estimated when sample size is large than when sample size is small. SIFASP–ML has poorer estimates at small sample sizes than the SCA–methods. However, the number of proper solutions on which the average DFC values are based is very small.

Because of the small number of cases in the present experiment, no significance tests were performed. The results give some indication as to how the behavior of the methods differ at the different sample sizes, but these results are not solid enough to warrant strong conclusions. However, it appears that at larger sample sizes, differences between the four methods are negligible.

Table 8.3: *DFC values for the proper solutions*

Sample Size	SIFASP–ML	SCA–W	SCA–P	SCA–S	number of proper solutions
50	.27	.11	.12	.17	3
100	.17	.08	.08	.09	2
150	.20	.06	.06	.07	6
300	.06	.04	.05	.05	9
500	.03	.03	.03	.03	10
1000	.03	.02	.02	.02	10

Answer 2: With larger sample sizes, SIFASP–ML seems better able to indicate the correct number of factors, in the PatternComb condition, used in the present experiment. However, the SCA–methods give a better recovery of factors. Even at a sample size of 500, SIFASP–ML still is not able to recover all factors correctly.

8.6.3 Experiment 4c: Overlapping factors

Question 3: Which of the methods has the highest scores on the success measures DI, RR and DFC, when factors within each group overlap?

In the present experiment, there was a total of 60 cases, and for the dimension indication the absolute number of correct indications is given instead of percentages. In PatternComb condition '$\mathbf{P}_r\mathbf{P}_r$' (30 cases), the measure QDA indicated the correct dimension 30 times for all SCA–methods. This is equal to results for PatternComb condition '$\mathbf{P}_a\mathbf{P}_a$' in Experiment 2. In PatternComb condition '$\mathbf{P}_s\mathbf{P}_s$' (30 cases), the measure QDA indicated the correct dimension 29 times for SCA–W and 28 times for SCA–P and SCA–S. This is one time less for each method than in PatternComb condition '$\mathbf{P}_b\mathbf{P}_b$' in Experiment 1. The measure ECVI indicated the correct dimension 56 times for SIFASP–ML; 29 times in PatternComb condition '$\mathbf{P}_r\mathbf{P}_r$' (this was 26 times in PatternComb condition '$\mathbf{P}_a\mathbf{P}_a$' in Experiment 2) and 27 times in PatternComb condition '$\mathbf{P}_s\mathbf{P}_s$' (no previous results).

For all models with the correct dimension, SIFASP–ML arrived at a proper solution, so no cases were left out for the success criteria RR and DFC. SCA–W and SCA–P had a RR of 99.2% (2 factors were not recovered correctly by each method in PatternComb condition '$\mathbf{P}_s\mathbf{P}_s$') and SCA–S and SIFASP–ML had a RR of 100%. In PatternComb condition '$\mathbf{P}_a\mathbf{P}_a$' in Experiment 2, all methods also had a RR of 100%, and in PatternComb condition '$\mathbf{P}_b\mathbf{P}_b$', this also was the case for the SCA–methods (no results for SIFASP–ML were available in this condition). Mean absolute DFC values over the 60 cases, and for the two PatternComb conditions in the present experiment and the two comparable PatternComb conditions from previous experiments are presented in Table 8.4. The SCA–methods always

Table 8.4: *Mean absolute DFC values*

	SIFASP–ML	SCA–W	SCA–P	SCA–S	cases
All cases	.26	.27	.27	.32	60
$\mathbf{P}_r\mathbf{P}_r$	.30	.32	.32	.36	30
$\mathbf{P}_a\mathbf{P}_a$ (exp.2)	.11	.10	.10	.10	30
$\mathbf{P}_s\mathbf{P}_s$	.21	.23	.23	.27	30
$\mathbf{P}_b\mathbf{P}_b$ (exp.1)	–	.09	.10	.11	30

overestimated the correlations, while SIFASP–ML showed no such bias. As can be seen from Table 8.4, the overestimations of the correlations are significantly larger than in the comparable conditions from previous experiments ($p<.001$, for all 2–tailed t–tests), demonstrating that the overlap causes higher estimates of the correlations. The differences between the methods on all three success criteria were very small.

Answer 3: When factors overlap, all methods are good at dimension indication and recovering factors, but the overlap causes the SCA–methods to severely overestimate (in absolute sense) the correlations between factors.

8.7 CONCLUSION

In the extreme situations with four groups, or two groups and sample sizes of 300 or more, the differences between the methods, that were present in a similar two group condition, and at smaller sample size, respectively, disappear. In the situation where there is overlap between factors, the dimension indicators used for the methods have little trouble indicating the correct number of factors and recovery of factors is close or equal to 100%. When factors overlap, however, all methods severely overestimate the absolute value of the correlations between factors.

CHAPTER 9

CONCLUSION AND DISCUSSION

In the present chapter, first an overview of the main results from the experiments will be given (Section 9.1). This overview is ordered as follows. Firstly, a summary of the strong and weak points of each method is presented and it will be decided which method is the best overall method (Section 9.1.1). Secondly, the qualities, specialties and flaws, of each method will be discussed in detail (Section 9.1.2). Thirdly, some of the results for the four dimension indicators, used for the SCA–methods, and the four dimension indicators, used for SIFASP–ML, will be summarized, and a suggestion for an alternative way of determining the number of components to retain will be presented (Section 9.1.3). In Section 9.1.4, the discriminant function, derived in Experiment 1, is reviewed and its performance on the data simulated in Experiments 3 and 4 is presented. After the overview, the limitations of this study will be discussed, focusing on the range of values used for the independent variables (Section 9.2.1), the manner of data generation (Section 9.2.2), and the use of the methods with a specific rotation procedure in this study (Section 9.2.3). Also, attention will be paid to the user friendliness and availability of the methods, investigated in this study (Section 9.3). Finally, in Section 9.4, some guidelines to be used when analyzing data sets from two or more groups, with scores on the same variables, will be presented.

9.1 OVERVIEW OF THE RESULTS AND DISCUSSION

9.1.1 The preferable method: SCA-P

In Table 9.1, an overview is presented of the performance of the

methods investigated in the present study, on the various measures used. The results will now be discussed for each method separately.

The results from the present experiments, as summarized in Table 9.1, have made it clear that there is one method that outperforms the others: SCA–P. This method (including the rotation, described in Chapter 2) was most successful in indicating the correct number of factors, using the dimension indicator QDA (in fact it gave the best overall results for all dimension indicators). The only noteworthy exception was in the condition where there were strong and weak factors present: Here the measure ECVI for SIFASP–ML had a higher number of correct dimension indications. Furthermore, SCA–P had the highest overall recovery rate,

Table 9.1: *Overview of the performance of each method on various success criteria*

	SCA–P	SCA–S	SCA–W	SIFASP–ML–C
Dimension Indication for				
– samples from one population	+/–	–	+/–	+
– samples from two populations	+	–	+	+/–
Recovery of factors for				
– samples from one population	+	+	+/–	+/–
– samples from two populations	+	–	+/–	+/–
– small samples (n=50)	+	+/–	–	+/–
– large samples (n=150 or more)	+	+	+	+
Recovery of factor correlations for				
– samples from one population	+	+	+	+/–
– samples from two populations	+	–	+	+
Recovery of relative factor strength	+/–	+	+/–	*n.r.a*
Performance in the presence of weak factors				
– in both groups	+	+/–	+/–	+/–
– in one group	+	+	–	*n.r.a*

n.r.a = no results were available in this condition. Note that the '+', '+/–' and '–' values in this table only indicate relative performance, so, for instance, a '–' does not necessarily mean that performance is bad

and (together with SCA–W) had the best overall approximation of the correlations between factors. It did not give the best recovery of relative factor strengths, but the superiority of SCA–S is offset by the fact that in some conditions, the factors were recovered less well than by SCA–P. An interesting characteristic of SCA–P was that strong factors were always correctly recovered within each sample, but this sometimes led to a lower recovery of weak factors present. Besides these recovery features, SCA–P also has the shortest execution time of the methods tested, making it the preferable method.

Before discussing flaws and assets of each method separately, it should be emphasized that, for samples from one population, in conditions where there were four groups, or two groups and sample sizes of 300 or more, hardly any differences between the methods appeared. In the situation where there was overlap between factors, recovery of factors was very high, but the absolute value of the correlations between factors was severely overestimated by all methods.

9.1.2 Characteristics of the methods SCA–S, SCA–W and SIFASP–ML

For each of the three other methods investigated except SCA–P, serious shortcomings have been detected and illustrated in the present study. The problems encountered will now be summarized for each of the other three methods separately. Also, for each method, some of its assets will be summarized. First, SCA–S will be reviewed, then SCA–W, and finally SIFASP–ML.

For SCA–S, the QA measure had the highest overall number of correct dimension indications, which was, unfortunately, not very high, compared to the other methods. The (relative) success of the measure QA was due to its performance in the situation where samples came from two (different) populations. When samples came from one population, the measure PA had the highest number of correct dimension indications, followed by the measure QDA and only in third position by the measure QA. The conclusion from these results is that for SCA–S the dimension indicators do not work well. SCA–S is just as good as SCA–P at recovering the underlying factors

and factor correlations when samples are coming from the same population, but is outperformed by SCA–P on both measures when the samples come from two distinctly different populations. The presence of weak factors decreases the recovery of strong factors. The only advantage of SCA–S over SCA–P is that it is most successful at recovering the relative strength of factors. This result is offset, however, by the fact that in the condition when the samples come from two distinctly different populations (i.e., do not have an equal or columnwise proportional underlying structure matrix) it relatively often fails to recover the factors. Hence the good recovery of the relative factor strength may quite often pertain to wrong components.

On each of the success criteria, SCA–W gave results that were equal to or worse (though usually not much) than those of SCA–P. Thus, the present study suggests that SCA–W has virtually nothing interesting to add to results obtained by SCA–P. Besides this, a serious flaw of SCA–W is that it gives relatively bad results when weak factors are present. The weak factors are often not recovered by the method, especially when sample size is small (n=50). Also, the presence of weak factors hampers the recovery of strong factors. The only virtue in the latter situation is that it is best in recovering the weak factors, but the gain compared to SCA–P is not large.

The present study suggests that SIFASP–ML should be used on covariances and not on correlations. The preferable dimension indicator for SIFASP–ML is the measure ECVI. It gave a higher or approximately equal number of correct dimension indications than the other dimension indicators for SIFASP–ML in all experiments. In situations where there are strong factors present, recovery of factors is high for SIFASP–ML. In situations where factors are weakly defined or where factor loadings are of different sizes across groups, SIFASP–ML may have trouble in arriving at a proper solution, even when the correct dimension is chosen. When looking at the cases for which a proper solution was found, overall recovery of factors, using SIFASP–ML, was high, but lower than when using SCA–P. SIFASP–ML had more trouble in recovering the factors than SCA–P when the number of variables is small (12 instead of 24) and when the

number of factors is large (4 instead of 2). It is also in these conditions that SIFASP–ML tends to have difficulty in arriving at a proper solution. This is the situation where there is a low variable–to–factor ratio. SIFASP–ML is better at recovering correlations of .4 between factors, than recovering zero correlations, but in both cases worse than SCA–P. Overall, SIFASP–ML gives somewhat higher DFC values than SCA–P, but differences between the methods are small.

9.1.3 Results for the dimension indicators

When applying SCA–W, SCA–P and PCA–sep, the measure QDA is the dimension indicator that can best be used. When applying SCA–S, the measure QA is the dimension indicator that can best be used. Surprisingly, differences between the dimension indicators are not much influenced by sample size, although all dimension indicators have higher success rates for larger sample sizes. The effect of the number of factors present is much stronger. Four components are much harder to be correctly indicated by the measures PA, QA and QDA, while the measure KA1 has more trouble correctly indicating two components. When four factors are present, all four dimension indicators perform about equally well for all methods, although the measure QDA still has the highest success rate.

The values of the QDA measure should be inspected visually instead of letting the computer pick the dimension for which the highest QDA values is found. In some experiments, for instance, where there were two strong factors and two weak factors present, the QDA measure attained its highest value for two components, which was judged as incorrect. However, the QDA measure also attained a very large value for four components, indicating that the choice between two components and four components is not at all clear–cut. This information, however, has been ignored completely in the present study.

For SIFASP–ML solutions, the measure ECVI can best be used for dimension indication, because this measure had the highest percentage of correct dimension indications of the fit indices tested. In samples from one population, the number of correct dimension indications for SIFASP–ML

(applied to covariances and using the measure ECVI) was about equal to or higher than that of SCA–P (using the measure QDA). Specifically, it was higher in PatternComb condition '$\mathbf{P}_e\mathbf{P}_e$', the condition where there are strong and weak factors in both groups. In samples from two (different) populations, for SIFASP–ML, the number of correct dimension indications was lower than for SCA–P.

Velicer and Jackson (1990) suggested that the failure of common factor analysis to arrive at a proper solution should be used as a diagnostic. In the present study, where SIFASP–ML was used as a generalization of common factor analysis, this property (failure to arrive at a proper solution) was frequently encountered, and in fact, it *was* used as a diagnostic, in the sense that it was not tried to fix the improper solutions, nor were they ignored, but, instead, the improper solutions were interpreted as a signal that the model which SIFASP–ML tried to fit, was simply not suited for the data set at hand and therefore should be abandoned. However, this use of the improper solutions as a diagnostic also frequently led to discarding the solution with the correct dimension.

While, in the one group situation, only common factor analysis uses an iterative procedure, in the two or more groups situation this is no longer the case. Some of the generalizations of component analysis, namely SCA–W and SCA–S, also make use of iterative procedures. In Section 4.6, it was investigated what convergence criterion could best be used in order to be relatively certain that the global optimum had been reached. As a side result of this experiment, it was found that, when the correct dimension was fed into the program, the global optimum of the minimized function was attained for both the rational and all of the random starts, while at an incorrect dimension, the global optimum of the minimized function was only attained for a relatively small number of different starting positions used. Viewed as a diagnostic, the property of not attaining the global optimum all of the time, was an indication that the wrong dimension was taken. Thus, the generalizations of component analysis using iterative procedures may very well possess a valuable additional instrument for determining the correct dimension. This subject

requires further study, but initial results look promising.

9.1.4 The discriminant function for discerning samples from one population from samples from two populations

In Section 5.6.5, a discriminant function was derived for distinguishing samples from one population from samples from two (or more) different populations. In Chapters 7 and 8, new data sets have been created that have now been used for a second validation of the discriminant function. For the data simulated in Experiments 3a to 4c, the results of applying function (5.1) are presented in Table 9.2.

From Table 9.2, it can be seen that the discriminant function succeeds in correctly classifying most of the data sets, with the only exception of all data sets in Experiment 4b, with sample sizes of 300 or more. These were all misclassified as samples from two different populations. This failure of the discriminant function at large sample sizes is not very surprising, however, as can be seen by looking at the discriminant function (5.1). In the discriminant function, there are four predictor variables: EV_S, EV_A (explained variances by SCA–S and PCA–sep, respectively), NVar (number of variables) and SS (sample size). Consider a situation in which there is a fixed number of variables present, and there are samples drawn, increasing in sample size. For the lower sample size, the discriminant function can be expected to correctly classify most samples. However, at larger sample sizes, the negative weight for sample size in the discriminant function will make sure that at some

Table 9.2: *Percentage of correct classifications of the samples as coming from 'one population' or as coming from 'two populations'*

Exp. 3a	Exp. 3b	Exp. 4a 1 pop.	Exp. 4a 2 pop.	Exp. 4b $n \leq 150$	Exp. 4b $n \geq 300$	Exp.4c
96.5%	82.5%	98%	88%	100.0%	0.0%	100.0%

sample size, the discriminant function value will become negative as well, because (in this example) the number of variables does not change, while the explained variances of SCA–S and PCA–sep in the function will always have values between 0 and 100 (%), and in fact, will not change all that much with increasing sample size.

The conclusion from this finding is that the discriminant function should only be used when sample sizes are maximal 150. This limitation of the applicability of the discriminant function is, in our view, not troublesome, because at large sample sizes the effect of random or measurement error is expected to be small. That is, the population values will be approached, making the use of a discriminant function to distinguish samples from one or two populations unnecessary, because the results from the analysis itself are reliable enough to base that judgment on. That is, when the solutions from SCA–P and SCA–S lead to the same interpretation of the components, the samples are from one population, and when the solutions from SCA–P and SCA–S lead to different interpretations of the components, the samples are from two (or more) populations.

9.2 LIMITATIONS OF THE PRESENT STUDY

In the present study, a number of choices had to be made that necessarily limited the range and generalizability of the results. In this section, three different kinds of choices will be addressed separately.

9.2.1 Beyond the values of the independent variables

The foremost limitation has been the choice of the values of the independent variables, used in the present experiment. These were chosen so as to cover the situations one is most likely to encounter in empirically gathered data. It seemed pointless to see how the methods under investigation would perform in situations that either never happen

or for which the methods are not meant. The underlying patterns and factor correlations used, cover such a broad range of possible situations, that generalization of the results seems warranted. For the independent variable 'sample size', it was found that at the smallest sample size used (n=50), the differences between the methods were the largest, and that already at a sample size of 150, differences between the methods had shrunk considerably. For larger sample size, differences between the methods disappeared (as was shown in Experiment 4b) when the samples came from one population. When samples come from different populations, this is bound to be reflected by different results found for SCA–P (or SCA–W) and SCA–S, especially. The values used for the independent variables 'number of variables' and 'number of factors' also reflect the situations one might encounter most often in empirical research, although a situation with a considerably larger number of variables is not uncommon. In that situation, the methods are likely to give better results, because it is expected that with the number of variables, the variable to factor ratio will increase as well. A situation where there are, for instance, 96 variables and 16 factors, is far removed from an empirical situation, and therefore, this kind of conditions was not included in the present study.

9.2.2 Data simulation

A serious limitation of our study, and with it, most of the simulated research done on the comparison of factor and component analysis, is the way the data are simulated. The assumption in this kind of simulation research is that the simulated data are similar to data gathered in empirical situations, but this is by no means always the case. The distribution of scores around underlying values in the data simulation is normal, but skewed distributions have been found regularly in practice.

An alternative approach to the one we have adopted is defining an underlying structure matrix and underlying factor correlations and simulating data based on these two matrices. With this approach, it is

possible to simulate data for two or more groups so, that all the simulated groups have the same underlying structure matrix, but a different underlying pattern matrix. For this kind of data, different results can be expected in the performance of the methods. However, because this kind of data has never been encountered in the literature on simulation studies, it was chosen not to include it in this study.

9.2.3 Methods and rotations

In the present study, for each method a rotational method was chosen to optimally simplify the solution and make interpretation possible. It is believed that with the rotations used, on the whole the underlying pattern(s) or structure(s) could be approximated as well as possible. It is conceivable, however, that in certain conditions, or even in separate data sets, a different rotational method would have given a better approximation of the underlying characteristics used in the data simulation.

9.3 USER FRIENDLINESS OF THE PROGRAMS USED

At this point, it seems appropriate to discuss the user–friendliness of the programs, used in this study. For SIFASP–ML, we used the commercial LISREL program (LISREL8e, Jöreskog and Sörbom, 1993). For SCA–W, there also exists a commercial program (called SCA, Kiers, 1990), but this was not used, because automation of the experiments was easier to realize by implementing the algorithm in MATLAB (Mathworks Inc.). This implementation of SCA–W gave exactly the same results as the SCA–program by Kiers. SCA–P has been implemented as an option in the SCA–program, although it is not named as such. Again, a newly implemented version of SCA–P in MATLAB was used. For SCA–S, there is no commercial version available and therefore this method was also implemented in MATLAB. For these reasons, the main focus of this section will be on the user–friendliness of the program LISREL8e, as used for SIFASP–ML.

For the analysis of more than one group, LISREL8e led to various disappointments, in that a number of the promised assets were not delivered by the program, without proper warnings of some kind. Although the problems, mentioned next, have no implications on the quality of SIFASP-ML – after all, the LISREL8e program is only one implementation of the method and not the method itself – we feel it is instructive to reveal some of the ill-constructed features of the program to warn future users for possible mishaps and disappointments.

The LISREL8 program (not the extended version LISREL8e) failed to handle analyses with as few as 24 variables. Both the programs LISREL8 and LISREL8e encountered situations where the initial estimates (TSLS for ML-estimates) produced a fitted covariance matrix that was not positive definite, leading to a fatal error. Next, in the two groups situation, the values of the goodness of fit indices GFI (Goodness of Fit Index) and the PGFI (Parsimony Goodness of Fit Index) were found to be incorrectly (incompletely) presented by the LISREL program. The GFI value was only presented for the last group in the analysis, and the PGFI was calculated for this incomplete GFI value, using the degrees of freedom of the complete model, leading to a PGFI value larger than one (the measure is defined to be between 0 and 1). Furthermore, the AGFI value was not presented at all in the multiple group situation.

A more serious flaw was discovered when it was tried to solve the problem of the measures GFI and PGFI by running the analysis twice, offering the samples in opposite order in each analysis, thus hoping to get two GFI and PGFI-values, from which the correct GFI and PGFI-values could be calculated. However, from the initial runs it appeared that the two solutions often (not always) differed, in that for all goodness-of-fit indices different values were obtained. Apparently, the program often converged to different solutions, when the samples were offered in different orders. For this reason, all attempts to acquire the correct GFI and PGFI-values were abandoned.

The program for SCA-W is much more user-friendly. This is mostly due to the fact that SCA-W needs little specification beforehand and there is no need for identification constraints. One simply enters the matrices to

be analyzed, and interactively inputs sample size, accuracy (=convergence criterion, optional choice) and the number of components to retain, among others. Furthermore, the program has several built–in rotation methods from which the user may choose to rotate the solution. As already said previously, however, we do not recommend the use of SCA–W.

SCA–P is built–in into the program for SCA–W (Kiers, 1990). Fortunately, all the assets of the program also hold when using the SCA–P option of the SCA–program. The only problem with this implementation of SCA–P is that the SCA–P option is hard to recognize. For SCA–S, there is no commercial program available.

9.4 GUIDELINES WHEN ANALYZING TWO OR MORE GROUPS

It may be clear from the results, presented above, that we advise *against* the use of SCA–W and SIFASP–ML and *in favor of* the use of SCA–P. Because of its additional features, we *also* advise to use SCA–S, complementary to the use of SCA–P. It is an unfortunate situation that the preferable methods are the ones that are either somewhat hidden in the SCA–program (SCA–P) or not available at all (SCA–S). However, this situation will hopefully change in the (near) future. Incorporating the results from the various conditions, investigated in this study, our guidelines for any analysis of two or more groups are the following:

1.) Start by calculating the eigenvalues of the sum of the correlation matrices for the samples to be analyzed (weighted by sample size). Determine the number of components to retain by means of the measure QDA for SCA–P and, if possible, compare this with the number of components expected. A smaller number of components is likely to be better indicated than a larger number of components, especially when there are some weak components present. Because the measure QDA will hardly ever overestimate the number of components, one should sooner consider adding an extra component than deleting one. The dimension indication gains accuracy with the number of variables. In case of

serious discrepancy of the number of components indicated by the measure QDA and the number of components expected on the basis of a theoretical model (in our view, the seriousness in the discrepancy is something that can only be determined by the researcher), we advise to inspect the values of the QDA measure for the various dimensions. It may be the case that the QDA measure attains a high value for two different numbers of dimensions, opting for two different solutions. In the present study, this was found when there were two strong factors and two weak factors present. In that condition, the measure QDA attained high values for both two components and four components. We ignored this information in the present study, but it is important to realize that a property of the measure QDA is that it will attain a large value for the correct dimension, although it may attain a larger value for an incorrect dimension. Theoretical expertise will have to be the final judge in a situation where the measure QDA behaves this way.

2.) After deciding on the number of components, next, do the analysis of the two or more groups for the thus derived dimension twice, firstly with SCA–P and secondly with SCA–S. Rotate the solution of SCA–P with the Harris and Kaiser's (1964) Independent Cluster rotation and the solution of SCA–S with the simple structure rotation, devised by Kiers and Ten Berge (1994a). Try interpreting the components, based on the pattern matrix, found with SCA–P, and on the basis of the structure matrix, found with SCA–S, or use the congruence measure to see whether the same components are represented by the two matrices. Do the components receive the same interpretation, when using the results from the two methods? If the sample sizes used are not larger than 150, the discriminant function (5.1) can be used as a complement to get an indication of whether the samples are likely to be from one or two (distinctly different) populations. A positive value indicates that the samples are probably from the same population, while a negative value indicates that the samples are probably from two different populations. Finally, remember that the number of variables is not of great influence on the correct recovery of factors, and, when sample sizes of 500 or more are used, the two methods SCA–P and SCA–S will usually not differ in

interpretation of components, when the samples are indeed drawn from one population. So any great differences between component interpretation in that case must lead to the conclusion that the samples are not from the same population, and simultaneous component analysis is not appropriate in that situation.

3a.) If the populations seem to be the same, lay aside the results from SCA–S and continue with the results from SCA–P: Interpret the components on the basis of the pattern matrix. To inspect possible small differences between the groups, inspect the structure matrices for each group.

3b.) If there seem to be considerable differences between the populations from which the samples were drawn, use the structure matrices from SCA–P to interpret the components for each of the populations.

4.) Interpreting the correlations, found with the methods, is a hazardous task, because recovery of component correlations was not very accurate. As a rule of thumb, based on the average DFC values found in this study, it should be kept in mind that the underlying correlations between the components are within $\pm$ .15 of the correlations found for SCA–P, and within $\pm$ .25 for SCA–S.

For illustrational purposes, two empirical data sets were analyzed following the guidelines just presented. Results of these analyses are presented in Appendix D.

The present study has led us to recommend the oldest and most basic method, SCA–P, for general use. We propose to complement it with the often somewhat idiosyncratic SCA–S method, not because of its particularly good performance, but for its help in deciding how to interpret the SCA–P results. The two more or less standard methods (SCA–W and SIFASP) turned out to have little to improve upon SCA–P, and are often clearly outperformed by SCA–P.

Appendix A True pattern matrices (order 12×2 and 12×4) used in this study

$$P_a=\begin{pmatrix} .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \end{pmatrix}, \quad P_b=\begin{pmatrix} .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \end{pmatrix}, \quad P_c=\begin{pmatrix} .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \end{pmatrix}, \quad P_d=\begin{pmatrix} .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \end{pmatrix},$$

$$P_e=\begin{pmatrix} .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \end{pmatrix}, \quad P_f=\begin{pmatrix} .37 & 0 \\ .55 & 0 \\ .89 & 0 \\ .37 & 0 \\ .55 & 0 \\ .89 & 0 \\ 0 & .37 \\ 0 & .55 \\ 0 & .89 \\ 0 & .37 \\ 0 & .55 \\ 0 & .89 \end{pmatrix}, \quad P_g=\begin{pmatrix} .89 & 0 \\ .55 & 0 \\ .37 & 0 \\ .89 & 0 \\ .55 & 0 \\ .37 & 0 \\ 0 & .89 \\ 0 & .55 \\ 0 & .37 \\ 0 & .89 \\ 0 & .55 \\ 0 & .37 \end{pmatrix}, \quad P_h=\begin{pmatrix} .55 & 0 \\ .37 & 0 \\ .89 & 0 \\ .55 & 0 \\ .37 & 0 \\ .89 & 0 \\ 0 & .55 \\ 0 & .37 \\ 0 & .89 \\ 0 & .55 \\ 0 & .37 \\ 0 & .89 \end{pmatrix},$$

$$P_l=\begin{pmatrix} .37 & 0 \\ .89 & 0 \\ .55 & 0 \\ .37 & 0 \\ .89 & 0 \\ .55 & 0 \\ 0 & .37 \\ 0 & .89 \\ 0 & .55 \\ 0 & .37 \\ 0 & .89 \\ 0 & .55 \end{pmatrix}, \quad P_m=\begin{pmatrix} .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \end{pmatrix}, \quad P_n=\begin{pmatrix} .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \end{pmatrix}, \quad P_p=\begin{pmatrix} .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \\ 0 & .37 \end{pmatrix},$$

$$P_q=\begin{pmatrix} .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ .37 & 0 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \end{pmatrix}, \quad P_r=\begin{pmatrix} .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .89 & 0 \\ .63 & .63 \\ .63 & .63 \\ .63 & .63 \\ .63 & .63 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \\ 0 & .89 \end{pmatrix}, \quad P_s=\begin{pmatrix} .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .55 & 0 \\ .39 & .39 \\ .39 & .39 \\ .39 & .39 \\ .39 & .39 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \\ 0 & .55 \end{pmatrix}.$$

$$P_a=\begin{pmatrix} .89 & 0 & 0 & 0 \\ .89 & 0 & 0 & 0 \\ .89 & 0 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & 0 & .89 \\ 0 & 0 & 0 & .89 \\ 0 & 0 & 0 & .89 \end{pmatrix}, \quad P_b=\begin{pmatrix} .55 & 0 & 0 & 0 \\ .55 & 0 & 0 & 0 \\ .55 & 0 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & 0 & .55 \\ 0 & 0 & 0 & .55 \\ 0 & 0 & 0 & .55 \end{pmatrix}, \quad P_c=\begin{pmatrix} .37 & 0 & 0 & 0 \\ .37 & 0 & 0 & 0 \\ .37 & 0 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & 0 & .37 \\ 0 & 0 & 0 & .37 \\ 0 & 0 & 0 & .37 \end{pmatrix},$$

$$P_d=\begin{pmatrix} .89 & 0 & 0 & 0 \\ .89 & 0 & 0 & 0 \\ .89 & 0 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & 0 & .55 \\ 0 & 0 & 0 & .55 \\ 0 & 0 & 0 & .55 \end{pmatrix}, \quad P_e=\begin{pmatrix} .89 & 0 & 0 & 0 \\ .89 & 0 & 0 & 0 \\ .89 & 0 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & 0 & .37 \\ 0 & 0 & 0 & .37 \\ 0 & 0 & 0 & .37 \end{pmatrix}, \quad P_f=\begin{pmatrix} .37 & 0 & 0 & 0 \\ .55 & 0 & 0 & 0 \\ .89 & 0 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & 0 & .37 \\ 0 & 0 & 0 & .55 \\ 0 & 0 & 0 & .89 \end{pmatrix},$$

$$P_g=\begin{pmatrix} .89 & 0 & 0 & 0 \\ .55 & 0 & 0 & 0 \\ .37 & 0 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & 0 & .89 \\ 0 & 0 & 0 & .55 \\ 0 & 0 & 0 & .37 \end{pmatrix}, \quad P_h=\begin{pmatrix} .55 & 0 & 0 & 0 \\ .37 & 0 & 0 & 0 \\ .89 & 0 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & 0 & .55 \\ 0 & 0 & 0 & .37 \\ 0 & 0 & 0 & .89 \end{pmatrix}, \quad P_l=\begin{pmatrix} .37 & 0 & 0 & 0 \\ .89 & 0 & 0 & 0 \\ .55 & 0 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & .89 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & 0 & .37 \\ 0 & 0 & 0 & .89 \\ 0 & 0 & 0 & .55 \end{pmatrix},$$

$$P_m=\begin{pmatrix} .37 & 0 & 0 & 0 \\ .37 & 0 & 0 & 0 \\ .37 & 0 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & 0 & .89 \\ 0 & 0 & 0 & .89 \\ 0 & 0 & 0 & .89 \end{pmatrix}, \quad P_n=\begin{pmatrix} .55 & 0 & 0 & 0 \\ .55 & 0 & 0 & 0 \\ .55 & 0 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & .89 & 0 \\ 0 & 0 & 0 & .89 \\ 0 & 0 & 0 & .89 \\ 0 & 0 & 0 & .89 \end{pmatrix}, \quad P_p=\begin{pmatrix} .55 & 0 & 0 & 0 \\ .55 & 0 & 0 & 0 \\ .55 & 0 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & .37 & 0 \\ 0 & 0 & 0 & .37 \\ 0 & 0 & 0 & .37 \\ 0 & 0 & 0 & .37 \end{pmatrix},$$

$$P_q=\begin{pmatrix} .37 & 0 & 0 & 0 \\ .37 & 0 & 0 & 0 \\ .37 & 0 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & .37 & 0 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & .55 & 0 \\ 0 & 0 & 0 & .55 \\ 0 & 0 & 0 & .55 \\ 0 & 0 & 0 & .55 \end{pmatrix}.$$

Appendix B Identification constraints used for the LISREL program

Constrained values for the matrix Λ, used in the conditions where there were 12 variables for the six different models (with one to six factors), as specified for the method SIFASP–ML for the LISREL8 program. All zeros and ones stand for fixed loadings (fixed at zero or one). These values could not change during the iterations of the program. All nonfixed elements of Λ were left completely free. For Experiment 4c, the constrained values were somewhat adjusted (see below).

$$\Lambda_1=\begin{pmatrix} 1 \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \end{pmatrix},\quad \Lambda_2=\begin{pmatrix} 1 & 0 \\ . & . \\ . & . \\ . & . \\ . & . \\ . & . \\ 0 & 1 \\ . & . \\ . & . \\ . & . \\ . & . \\ . & . \end{pmatrix},\quad \Lambda_3=\begin{pmatrix} 1 & 0 & 0 \\ . & . & . \\ . & . & . \\ . & . & . \\ 0 & 1 & 0 \\ . & . & . \\ . & . & . \\ . & . & . \\ 0 & 0 & 1 \\ . & . & . \\ . & . & . \\ . & . & . \end{pmatrix},\quad \Lambda_4=\begin{pmatrix} 1 & 0 & 0 & 0 \\ . & . & . & . \\ . & . & . & . \\ 0 & 1 & 0 & 0 \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & 1 & 0 \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & 0 & 1 \\ . & . & . & . \\ . & . & . & . \end{pmatrix},$$

$$\Lambda_5=\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 1 & 0 & 0 & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 1 & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & 1 & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & 0 & 1 \\ . & . & . & . & . \end{pmatrix},\quad \Lambda_6=\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 1 & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 1 & 0 & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 1 & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 1 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 1 \\ . & . & . & . & . & . \end{pmatrix}.$$

In the condition of 24 variables, the first 12 rows of the matrix Λ with starting values looked the same as above. The matrix was completed by stacking a matrix with 12 rows in which no values were fixed below this matrix.

For Experiment 4c, the constrained values, as presented below, were used.

$$\Lambda_1 = \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, \quad \Lambda_2 = \begin{pmatrix} 1 & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 & 1 \end{pmatrix}, \quad \Lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 1 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}, \quad \Lambda_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\Lambda_5 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \quad \Lambda_6 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

Appendix C Deriving the largest possible difference in explained variance between the methods SCA-W and SCA-P in case two rank q matrices are analyzed

Here, it will be shown that the largest possible difference in explained variance between the methods SCA-W and SCA-P, in the case where two rank q correlation matrices of order $m \times m$ are analyzed, is 50% (see Section 3.3.3). This happens when the true pattern matrices $\mathbf{P}_{t1}$ and $\mathbf{P}_{t2}$ are chosen orthogonal. In this situation the components found with SCA-W explain 100% of the variance and the components found with SCA-P explain 50% of the variance. This can be proven as follows:

Define an $m \times q$ pattern matrix $\mathbf{P}_{t1}$ ($q \leq \frac{1}{2}m$), in which each component pertains to the same number of m/q variables, the q components do not overlap, and all nonzero loadings are $\pm$ 1. Choose a second pattern matrix $\mathbf{P}_{t2}$ in the same way, with the additional requirement that the columns of the matrices $\mathbf{P}_{t1}$ and $\mathbf{P}_{t2}$ are orthogonal. Choose the matrix $\mathbf{\Phi}$ as $\mathbf{I}$. The error free population correlation matrices are now written as $\mathbf{R}_{t1}=\mathbf{P}_{t1}\mathbf{P}'_{t1}$ and $\mathbf{R}_{t2}=\mathbf{P}_{t2}\mathbf{P}'_{t2}$. The total amount of variance to be explained is $2m$. In this situation, the q-dimensional solution, found with the method SCA-W, explains 100% of variance, as shown in Section 3.3.3. The method SCA-P, however, only explains 50% of the total variance as will now be shown. That SCA-P never explains less than 50% follows from the fact that q components can always explain 100% of variance in one of the two groups. The q-dimensional solution of the method SCA-P is found by taking the first q eigenvectors of $\mathbf{R}_{sum}$ ($=\mathbf{R}_{t1}+\mathbf{R}_{t2}=\mathbf{P}_{t1}\mathbf{P}'_{t1}+\mathbf{P}_{t2}\mathbf{P}'_{t2}$). Because the columns of the matrices $\mathbf{P}_{t1}$ and $\mathbf{P}_{t2}$ are orthogonal, the matrix $\mathbf{R}_{sum}$ has rank $2q$. Furthermore, $\mathbf{R}_{sum}=(\mathbf{P}_{t1}|\mathbf{P}_{t2})(\mathbf{P}_{t1}|\mathbf{P}_{t2})'$, hence $(\mathbf{P}_{t1}|\mathbf{P}_{t2})$ contains all eigenvectors of $\mathbf{R}_{sum}$ that correspond to nonzero eigenvalues. In fact, these eigenvalues are all m/q. Hence, taking $\mathbf{K}$ as any set of q normalized columns of $(\mathbf{P}_{t1}|\mathbf{P}_{t2})$ maximizes $\mathrm{tr}(\mathbf{K}'\mathbf{R}_{sum}\mathbf{K})$, and any such set of columns gives the best SCA-P solution. Each such component corresponds to an eigenvalue of m/q, hence together these q components explain $\frac{m}{2m} * 100\% = 50\%$ of the variance.

Appendix D Examining real life data

In this appendix, the suggestions for practical use of SCA–P and SCA–S, made at the end of Chapter 9, will be illustrated on two empirical data sets. These have been analyzed previously, but not with SCA–P and SCA–S.

D.1 MEREDITH (1964) DATA

The first data set used was the same as the one analyzed by Meredith (1964). Meredith used a selection of tests from a monograph by Holzinger and Swineford (1939). The purpose of Meredith's analysis was to illustrate the performance of two procedures for rotating any number of factor pattern matrices based on different populations to conform to a single "best fitting" factor pattern matrix. Here, on the other hand, this data set is used to illustrate the SCA–methods. The data consist of scores of seventh and eight grade students from two schools on nine tests. The two schools had a different socioeconomic character. One (the Pasteur school) enrolling children of factory workers and one (the Grant–White school) enrolling children in a middle–class suburban area. In the selection of nine tests used by Meredith three different types of tests were administered. They were chosen so that there were a "space" factor, a verbal factor, and a memory factor present, each factor represented by three tests. These tests were: A Visual Perception test involving logical sequences of abstract figures, a Cubes test with designs on the surfaces, and a Paper Form Board test for the "space" factor; a General Information test, a Sentence Completion test, and a Word Classification test for the verbal factor; and a Figure recognition test, an Object–Number paired associates test, and a Number–Figure paired associates test for the memory factor. A complete description of the samples and tests can be found in Holzinger and Swineford (1939).

Meredith (1964) split the two groups of children into two groups each, by splitting at the median (within each school) on one of the speeded tests (an addition test; in the selection of nine tests by

Meredith, speeded tests were not included). This yielded four groups, denoted by Pasteur–Low (speed), Pasteur–High, Grant–White–Low, and Grant–White–High. The four groups had sample sizes of 77, 79, 74 and 71, respectively.

Data from the four groups were first analyzed with SCA–P. That is, the eigenvalues of the sum of the weighted correlation matrices were calculated and the values of the QDA measure were derived from these (see Section 4.2.4) to determine the number of components to retain. Only for the one, two and three–dimensional solution, the QDA value was acceptable (only for those three solutions the additional amount of explained variance was larger than the amount of variance explained by one variable). These three QDA values for the one, two and three–dimensional solution were 8.42, .45 and 5.42, respectively, indicating that one or three components should be retained. On theoretical grounds, three components were retained.

For both SCA–P and SCA–S, the three–dimensional solution was calculated. The solution found with SCA–P was rotated using the HKIC rotation (see Section 2.4.1); the solution found with SCA–S was rotated using the simple structure rotation described in Section 2.4.3. The pattern matrix, found with SCA–P, was normalized so that $\mathbf{P}'\mathbf{P}=\mathbf{I}$. Note that with the normalization the pattern equals the weights matrix. Also for the three–dimensional solution, the mean percentage of explained variance of four separate PCA's for the four groups was calculated. The percentage of explained variance for the three–dimensional solution was 64.3% for the separate PCA's, 62.9% for SCA–P and 60.9% for SCA–S. The difference between SCA–P and SCA–S is 2.0%, which is considered small, but not negligible, suggesting that there are small differences between the groups. The congruences of the columns of the pattern matrix found with SCA–P and the corresponding columns of the structure matrix found with SCA–S were .93, .89 and .89 for the three components, respectively. This suggests the samples were drawn from the same population. As group size for the discriminant function (5.1), the mean group size of the four groups 75.25 was chosen. The value of the discriminant function thus found was 1.32, also suggesting that the samples were drawn from the same

population. Besides inspecting the pattern matrix, found with SCA–P, also the structure matrices, found with SCA–P will be inspected, to detect possible small differences in behavior between the components in the four groups.

The loadings of the pattern matrix found with SCA–P are presented in Table D.1. As can be deduced from the loadings, the three components can be labeled as a "space" component (the ability to picture an object and perform actions (rotations) upon it), a verbal component, and a memory component. The relatively high loading of the test 'Figure Recognition' on the space component, besides the memory component, is not surprising, because this test clearly involves visual perception in addition to memory.

The structure matrix for each group is presented in Table D.2. As can be seen from Table D.2, the variables (tests) correlate highest with the component to which they pertain, with the exception of the variable 'Figure Recognition' in groups 1 and 4, that has a higher correlation with the 'space' component in those groups.

The most striking difference between the two Pasteur groups and the two Grant–White groups is that for the latter the components show much more overlap than for the Pasteur groups. This can also be seen in the correlations between the components, which are .19 when averaged over the

Table D.1: *Pattern matrix for the three dimensional solution, found with the method SCA–P*

	space	verbal	memory
Visual Perception	.49	.10	.10
Cubes	.58	–.07	–.10
Paper Form Board	.54	.02	–.11
General Information	.01	.58	–.05
Sentence Completion	–.05	.61	–.01
Word Classification	.05	.52	.09
Figure Recognition	.34	–.06	.39
Object–Number	–.14	.02	.67
Number–Figure	.05	–.02	.60

two Pasteur groups and .38 when averaged over the two Grant–White groups. As a consequence, the six components in the two Grant–White groups are all stronger than the corresponding components in the Pasteur groups. When comparing the Low–speed groups, we see that the differences in component strengths are considerable for both the verbal component and the memory component. When comparing the High–speed groups, the differences in component strengths are considerable for both the space component and the memory component. These findings can all be attributed

Table D.2: *Variable – component correlations for the three dimensional solution, found by the method SCA–P, in the four groups, and the component strengths, in percentages of explained variance*

	Group 1: Pasteur–Low			Group 2: Pasteur–High		
	space	verbal	memory	space	verbal	memory
Visual Perception	.77	.38	.32	.70	.43	.15
Cubes	.73	.06	.04	.67	.12	–.15
Paper Form Board	.74	.08	.02	.64	.22	–.22
General Information	.17	.90	–.01	.29	.87	.13
Sentence Completion	.11	.91	.12	.35	.90	.09
Word Classification	.38	.82	.36	.30	.84	.16
Figure Recognition	.59	.19	.56	.33	.09	.53
Object–Number	.06	.06	.81	–.12	.06	.77
Number–Figure	.11	.10	.76	–.00	.11	.72
Component strength	24.6%	28.0%	20.0%	19.7%	28.3%	17.1%
	Group 3: Grant–W–Low			Group 4: Grant–W–High		
	space	verbal	memory	space	verbal	memory
Visual Perception	.75	.49	.38	.75	.34	.47
Cubes	.70	.20	.14	.70	.19	.16
Paper Form Board	.70	.35	.12	.67	.36	.29
General Information	.42	.86	.34	.39	.86	.11
Sentence Completion	.33	.91	.35	.29	.88	.18
Word Classification	.41	.83	.43	.39	.83	.26
Figure Recognition	.59	.23	.64	.66	.24	.65
Object–Number	.09	.36	.81	.14	.16	.84
Number–Figure	.37	.32	.78	.51	.13	.78
Component strength	27.8%	32.7%	25.0%	28.7%	28.6%	24.1%

to the presence of a stronger first unstated component (largest eigenvalue) in the Grant–White groups (explaining 36.6% and 39.4% of the total variance, respectively) than in the Pasteur groups (32.4% and 30.3%, respectively).

D.2 VAN SCHUUR (1984) DATA

The second data set used was taken from Van Schuur (1984), and was previously used by Ten Berge (1986c) to illustrate the crossvalidation of weights, and by Ten Berge and Kiers (1990) and Kiers and Ten Berge (1994a) to give an example of the application of SCA–W. The data set consists of scores of two samples of party–activists (an Italian sample of 718 persons and a Danish sample of 1565 persons) who rated 15 political issues on Likert scales. The scale ran from 1 (very much against) to 5 (very much in favor). The issues are Fight inflation (INFL), Accelerate European Integration (EURO), Reduce Public Control (PUBL), Women Decide About Abortion (ABOR), Fight unemployment (EMPL), Defense Against Superpowers (SUPP), Punish Terrorists (TERR), Develop Nuclear Energy (NUCL), Control Multinationals (MULT), Protect Environment (ENVM), Reduce Regional Differences (REGI), Equal Opportunity for Men and Women (EQUA), Increase Military Expenditure (MILT), Reduce Income Differences (INCO), and Concern Own Needs Versus needs of Third World (THRD).

The two groups were first analyzed by SCA–P. It must be noted that, because of the large group sizes, it was decided not to weight the matrices of the two groups according to the sample size, as is described in Sections 2.2.2 and 2.5.2. Both group sizes were considered large enough to produce fairly accurate estimates of population values. In the previous SCA analyses of this data set, weighting according to the group sizes was also not used. The eigenvalues of the sum of the correlation matrices were calculated to determine the number of components to retain, using the QDA measure. Only for the one, two and three–dimensional solution, the QDA value was acceptable (only for those three solutions

the additional amount of explained variance was larger than the amount of variance explained by one variable). These three QDA values for the one, two and three–dimensional solution were 1.8, 13.0 and 14.4, respectively, indicating that two or three components should be retained. Although in the previous analyses of this data set, always two components were retained, in the present analysis, both the two and three–dimensional solutions were inspected.

For both SCA–P and SCA–S, first the two dimensional solution was calculated. The solutions, found with SCA–P and SCA–S, were obtained by the same rotations as in the previous section. Also for the two–dimensional solution, the mean percentage of explained variance of two separate PCA's for the two groups was calculated. The percentage of explained variance for the two–dimensional solution was 44.5% for the separate PCA's, 43.5% for SCA–P and 40.3% for SCA–S. The difference between SCA–P and SCA–S is 3.2%, which is considered not negligible. The congruences of the columns of the pattern matrix found with SCA–P and the corresponding columns of the structure matrix found with SCA–S were .97 and .91 for the two components, respectively. Because the sample sizes were very large, the fact that SCA–P and SCA–S lead to virtually the same interpretations of the components suggests that the samples were drawn from the same population. The loadings of the pattern matrix found with SCA–P, and the structure matrices for the two groups are presented in Table D.3.

The loadings (which equal the SCA–P weights), reported in the first two columns of Table D.3, differ very little from the weights found by Kiers and Ten Berge (1994a) with SCA–W. In fact, all SCA–P weights differ no more than .02 from the corresponding SCA–W weights reported by Kiers and Ten Berge (1994a), with the exception of the loadings for the Abortion issue, which differ .05 and .04 from the corresponding weights from SCA–W, for the two components, respectively. Therefore, the two components can be labeled the same as was done by Kiers and Ten Berge (1994a). The first component is thus labeled as a measure for 'Conservatism' and the second component as a measure for 'Progressiveness'. The correlations between the components are equal

(when rounded) to the ones found by Kiers and Ten Berge (1994a): respectively –.14 in the Italian sample and –.47 in the Danish sample. The highest difference in correlations between variables and components, compared to the solution found with SCA-W, is again found for the Abortion issue (still a negligible difference of .03). The main differences between the Danish and the Italian sample are that in the Danish sample the issues 'Control Multinationals' and 'Reduce Income Differences' are more strongly opposed by the conservatives than in the Italian sample and that the 'Women Decide About Abortion' issue is strongly opposed by the conservatives in the Italian sample, while this is not an issue for the conservatives in the Danish sample. For a more thorough inspection of the two dimensional solution, refer to Kiers and Ten Berge (1994a).

Table D.3: *Component pattern and structure matrices for the two dimensional solution, found by the method SCA–P, and the component strengths, in percentages of explained variance*

	Simultaneous Pattern		Structure in Danish sample		Structure in Italian sample	
component	1	2	1	2	1	2
INFL	.27	.27	.36	.18	.42	.42
EURO	.38	.06	.73	–.27	.69	.03
PUBL	.37	–.04	.70	–.43	.78	–.15
ABOR	–.17	.10	–.17	.22	–.58	.29
EMPL	.08	.36	–.03	.45	.00	.64
SUPP	–.08	.24	–.34	.46	–.20	.45
TERR	.41	.04	.72	–.25	.82	–.08
NUCL	.40	–.03	.78	–.40	.80	–.15
MULT	–.07	.38	–.52	.71	–.12	.63
ENVM	–.00	.41	–.38	.65	–.04	.71
REGI	.10	.43	–.07	.58	.04	.74
EQUA	–.08	.36	–.40	.69	–.27	.61
MILT	.38	–.05	.77	–.36	.73	–.23
INCO	–.15	.31	–.61	.64	–.26	.56
THRD	.30	.00	.62	–.33	.52	–.01
Percentage of explained variance			29.0	22.6	26.0	20.8

As a next step, for both SCA–P and SCA–S the three dimensional solution was calculated. The solutions, found with SCA–P and SCA–S, were again obtained in the same way as in the previous section. The percentage of explained variance for the three–dimensional solution was 52.5% for separate PCA's, 50.3% for SCA–P and 49.3% for SCA–S. The difference between SCA–P and SCA–S is now only 1.0% (compared to 3.2% for the two–dimensional solutions), which is considered negligible. The congruences of the columns of the pattern matrix found with SCA–P and the corresponding columns of the structure matrix found with SCA–S were .95, .84 and .67 for the three components, respectively. This suggested the samples were drawn from different populations. The loadings of the pattern matrix found with SCA–P, and the structure matrices for the two groups are presented in Table D.4.

From Table D.4 it can be seen that the loadings for the first component differ .06 or less from the loadings for the first component in the two–dimensional solution. The variable–component correlations in the two groups never differ more than .05 with the correlations found in the two–dimensional solution. Thus, the first component can again be labelled as a measure for 'Conservatism'.

The second component, however, is not equal to the second component from the two–dimensional solution. The component, labeled as a measure for 'Progressiveness' in the two–dimensional solution, is split into two components measuring different aspects of 'Progressiveness' in the three–dimensional solution. The first 'Progressiveness' component can be attached the label 'striving for equality of all individuals', as can be deduced from the large loadings (>.30) for the issues 'Control Multinationals', 'Reduce Regional Differences', 'Equal Opportunity for Men and Women' and 'Reduce Income Differences'. One could also label this component as a measure for 'Socialism'. Apparently, agreeing with these issues seems to coincide with support for the 'Protect the Environment' issue. The second 'Progressiveness' component is harder to interpret, because the large loadings (>.30) for the issues 'Women Decide About Abortion', 'Fight unemployment', 'Defense Against Superpowers' and 'Concern for Own Needs Versus needs of Third World' are not easily put

under a common denominator.

The two 'Progressiveness' components overlap considerably, as can be seen from their positive correlation (.30 in the Danish sample and .50 in the Italian sample) against negative correlations with the 'Conservatism' component (−.48 and −.21 in the Danish sample and −.05 and −.33 in the Italian sample for the first and second 'Progressiveness' components, respectively).

The main differences with respect to the structure matrices between the Danish and the Italian sample (besides the same differences on the 'Conservatism' component as in the two-dimensional solution) are that, for the 'Progressiveness and striving for equality of all individuals' component, the issues 'Reduce Public Control', 'Develop Nuclear Energy', 'Increase Military Expenditure' and 'Concern for Own Needs Versus needs

Table D.4: *Component pattern and structure matrices for the three dimensional solution, found by the method SCA−P, and the component strengths, in percentages of explained variance*

	Simultaneous Pattern			Structure in Danish sample			Structure in Italian sample		
component	1	2	3	1	2	3	1	2	3
INFL	.30	.17	.25	.39	.07	.29	.45	.40	.20
EURO	.37	.10	−.12	.72	−.26	−.29	.69	.07	−.27
PUBL	.38	−.09	.05	.71	−.50	−.10	.77	−.13	−.30
ABOR	−.11	−.14	.57	−.13	.05	.61	−.55	.16	.69
EMPL	.13	.19	.43	.01	.32	.53	.05	.56	.61
SUPP	−.02	.05	.50	−.30	.32	.64	−.16	.33	.64
TERR	.41	.01	.00	.73	−.30	−.11	.82	−.06	−.29
NUCL	.39	.01	−.15	.77	−.39	−.33	.79	−.10	−.42
MULT	−.06	.41	.03	−.51	.74	.27	−.08	.63	.38
ENVM	−.00	.46	−.04	−.36	.69	.19	−.01	.74	.31
REGI	.10	.50	−.07	−.07	.64	.02	.07	.77	.34
EQUA	−.06	.34	.13	−.38	.68	.37	−.24	.60	.44
MILT	.36	−.00	−.16	.76	−.33	−.38	.71	−.20	−.40
INCO	−.14	.36	−.02	−.60	.68	.24	−.24	.58	.33
THRD	.33	−.15	.30	.63	−.45	.10	.54	−.08	.08
Percentage of explained variance				28.5	23.0	12.1	25.6	19.6	16.9

of Third World' are more strongly opposed by the progressive party activists in the Danish sample than in the Italian sample and that the 'Fight unemployment' and the 'Reduce Income Differences' issues are more strongly favored by the progressive party activists in the Italian sample than by the progressive party activists in the Danish sample. The only interesting difference on the third component between the Danish and the Italian sample is that the issue 'Reduce Regional Differences' is of more importance for the progressive party activists in the Italian sample than in the Danish sample, probably because there *are* larger regional differences in Italy than in Denmark.

The two examples presented above demonstrate the application of the simultaneous components analysis method SCA-P in practice. Compared to separate analyses of the groups in a data set, the simultaneous analysis has several advantages. The common components found are easily interpreted using the pattern matrix (which after HKIC rotation is equal to the weights matrix), while possible (small) differences are discovered by looking at the structure matrices found for each group in the simultaneous analysis separately. Thus, the method SCA-P offers more insight into what is going on in a data set than performing separate component or factor analyses on the different groups. In the latter case, it is, for instance, possible that some of the components obtained from most groups are present in the other groups as well, but are too weak to emerge among the first principal components drawn from those groups (see Section 1.4). In SCA-P this problem is circumvented in an elegant way.

REFERENCES

Bentler, P.M. (1989). EQS structural equations program manual. *BMDP Statistics Software*. Los Angeles, California.

Bollen, K.A. (1989). *Structural equations with latent variables*. John Wiley & Sons, New York.

Brogden, H.E. (1969). Pattern, structure, and the interpretation of factors. *Psychological Bulletin*, 72, 375–378.

Browne, M.W., & Cudeck, R. (1989). Single sample cross–validation indices for covariance structures. *Multivariate Behavioral Research*, 24, 445–455.

Browne, M.W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21, 230–258.

Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N–way generalization of Eckart–Young decomposition. *Psychometrika*, 35, 283–319.

Carroll, J.D., & Chang, J.J. (1972). *IDIOSCAL (Individual Differences In Orientation SCALing): A generalization of INDSCAL allowing IDIOsyncratic reference systems as well as an analytic approximation to INDSCAL.* Paper presented at meeting of the Psychometric Society, Princeton, N.J., March 1972.

Carroll, J.D., Pruzansky, S., & Kruskal, J.B. (1980). CANDELINC: A general approach to multidimensional analysis of many–way arrays with linear constraints on parameters. *Psychometrika*, 45, 3–24.

Carroll, J.D., & Wish, M. (1974). Models and methods for three–way multidimensional scaling. In D.H. Krantz, R.C. Atkinson, R.D. Luce, & P. Suppes (Eds.): *Contemporary developments in mathematical psychology, Vol. II: Measurement, psychophysics, and neural information processing* (pp.57–105). San Francisco: Freeman.

Cliff, N. (1988). The eigenvalues–greater–than–one rule and the reliability of components. *Psychological Bulletin*, 103, 276–279.

Cudeck, R., & Browne, M.W. (1983). Cross–validation of covariance structures. *Multivariate Behavioral research*, 18, 147–167.

De Leeuw, J. & Heiser, W.J. (1982). Theory of multidimensional scaling. In P.R. Krishnaiah and L.N. Kanai (Eds.), *Handbook of Statistics*, *Vol.* 2 (pp.285–316). Amsterdam: North–Holland.

Dunn, T.R., & Harshman, R.A. (1982). A multidimensional scaling model for the size–weights illusion. *Psychometrika*, 47, 25–45.

Flury, B.H. (1984). Common principal components in k groups. *Journal of the American Statistical Association*, 79, 892–898.

Flury, B.H. (1987). Two generalizations of the common principal component model. *Biometrika*, 74, 59–69.

Flury, B.H. (1988). *Common principal components and related multivariate models*. John Wiley & Sons, New York.

Glorfeld, L.W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55, 377–393.

Gorsuch, R.L. (1983). *Factor analysis* (2nd. ed.). Hillsdale, N.J.: Lawrence Erlbaum.

Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–12.

Hakstian, R.A. (1976). Two-matrix orthogonal rotation procedures. *Psychometrika*, 41, 267–272.

Harris, C.W. & Kaiser, H.F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, 29, 347–362.

Harshman, R.A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an 'exploratory' multi-modal factor analysis. *UCLA working papers in Phonetics*, 16, 1–84.

Harshman, R.A. (1972). Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA working papers in Phonetics*, 22, 111–117.

Harshman, R.A., & Lundy, M.E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H.G. Law, C.W. Snyder, J.A. Hattie, & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pp. 122–215). New York,: Praeger.

Harshman, R.A., & Lundy, M.E. (1994). PARAFAC: Parallel factor analysis. *Computational Statistics and Data Analysis*, 18, 39–72.

Harshman, R.A., & Lundy, M.E. (1996). Uniqueness proof for a family of models sharing features of Tucker's three-mode factor analysis and PARAFAC-CANDECOMP. *Psychometrika*, 61, 133–154.

Haven, S., & Ten Berge, J.M.F. (1977). Tucker's coefficient of congruence as a measure of factorial invariance: An empirical study. *Heymans Bulletin* 290 EX. Department of Psychology, University of Groningen, Groningen.

Holzinger, K., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs*, No. 48. Chicago: Univ. Chicago Press.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185.

Horn, J.L., & Knapp, J.R. (1973). On the subjective character of the empirical base of Guilford's structure-of-intellect model. *Psychological Bulletin*, 80, 33–43.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.

Jöreskog, K.G., & Sörbom, D. (1993). *LISREL* 8: *User's reference guide*. Scientific software international, Lawrence Erlbaum Associates Publishers, Hillsdale, New Jersey.

Kaiser, H.F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187–200.

Kiers, H.A.L. (1990). *SCA: A program for simultaneous components analysis*. Groningen: IEC ProGamma.

Kiers, H.A.L. (1993). A comparison of techniques for finding components with simple structure. In C.M. Cuadras and C.R. Rao (Eds.),

Multivariate analysis: Future directions (pp. 67–86). Amsterdam: Elsevier Science Publishers.

Kiers, H.A.L., & Ten Berge, J.M.F. (1989). Alternating least squares algorithms for simultaneous components analysis with equal component weight matrices in two or more populations. *Psychometrika*, 54, 467–473.

Kiers, H.A.L. & Ten Berge, J.M.F. (1994a). Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure. *British Journal of Mathematical and Statistical Psychology*, 47, 109–126.

Kiers, H.A.L. & Ten Berge, J.M.F. (1994b). The Harris–Kaiser independent cluster rotation as a method for rotation to simple components weights. *Psychometrika*, 59, 81–90.

Kiers, H.A.L., Cléroux, R., & Ten Berge, J.M.F. (1994). Generalized canonical analysis based on optimizing matrix correlations and a relation with IDIOSCAL. *Computational Statistics and Data Analysis*, 18, 331–340.

Kroonenberg, P.M. (1983). *Three-mode principal component analysis: Theory and applications*. Leiden: DSWO Press.

Kroonenberg, P.M., & De Leeuw, J. (1980). Principal components analysis of three–mode data by means of alternating least squares algorithms. *Psychometrika*, 45, 69–97.

Levin, J. (1966). Simultaneous factor analysis of several Gramian matrices. *Psychometrika*, 31, 413–419.

McDonald, R.P. (1989). An index of goodness–of–fit based on noncentrality. *Journal of Classification*, 6, 97–103.

McDonald, R.P., & Mulaik, S.A. (1979). Determinacy of Common Factors: A nontechnical review. *Psychological Bulletin*, 86, 297–306.

Meredith, W. (1964). Rotation to achieve factorial invariance. *Psychometrika*, 29, 187–206.

Millsap, R.E., & Meredith, W. (1988). Component analysis in cross–sectional and longitudinal data. *Psychometrika*, 53, 123–134.

Mulaik, S.A., & McDonald, R.P. (1978). The effect of additional variables on factor indeterminacy in models with a single common factor. *Psychometrika*, 43, 177–192.

Norušis, M.J. (1990). *SPSS advanced user's guide*. Chicago: SPSS Inc.

Oud, J.H.L., Haughton, D.M.A., & Jansen, R.A.R.G. (1996). Information and other criteria in structural equation model selection. *Kwantitaieve Methoden*, 17, 69–100.

Snook, S.C., & Gorsuch, R.L. (1989) Component analysis versus common factor analysis: A Monte Carlo study. *Quantitative Methods In Psychology*, 106, 148–154.

Steiger, J.H. (1979). Factor indeterminacy in the 1930's and the 1970's: Some interesting parallels. *Psychometrika*, 44, 157–168.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.

Steiger, J.H., & Lind, J. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the meeting of the Psychometric Society, Iowa City, IA.

Steiger, J.H., & Schönemann, P.H. (1978). A history of factor

indeterminacy. In S. Shye (Ed.), *Theory construction and data analysis in the behavioral sciences* (pp.136–178). San Francisco: Jossey–Bass.

Steiger, J.H., Shapiro, A., & Browne, M.W. (1985). On the asymptotic distribution of sequential chi–square statistics. *Psychometrika*, 50, 253–264.

Ten Berge, J.M.F. (1977). Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, 42, 267–276.

Ten Berge, J.M.F. (1986a). Some relationships between descriptive comparisons of components from different studies. *Multivariate Behavioral Research*, 21, 29–40.

Ten Berge, J.M.F. (1986b). Rotation to perfect congruence and the cross–validation of component weights across populations. *Multivariate Behavioral Research*, 21, 41–64 & 262–266.

Ten Berge, J.M.F. (1986c). Rotatie naar perfecte congruentie en de Multipele Groep Methode. *Nederlands Tijdschrift voor de Psychologie*, 41, 218–225.

Ten Berge, J.M.F. (1988). Generalized approaches to the Maxbet problem and the Maxdiff problem, with applications to canonical correlations. *Psychometrika*, 53, 487–494.

Ten Berge, J.M.F. (1993). *Least squares optimalization in multivariate analysis*. Leiden: DSWO Press.

Ten Berge, J.M.F. (1994). When SCA–W explains an equal amount of variance as SCA–P, SCA–S will also explain that amount of variance. *Unpublished note.*

Ten Berge, J.M.F., Bekker, P.A., & Kiers, H.A.L. (1994). Some clarifications of the Tuckals2 algorithm applied to the IDIOSCAL problem. *Psychometrika*, 59, 193–201.

Ten Berge, J.M.F., Kiers, H.A.L., & Krijnen, W.P. (1993). Computational solutions for the problem of negative saliences and nonsymmetry in INDSCAL. *Journal of Classification*, 10, 115–124.

Ten Berge, J.M.F., & Kiers, H.A.L. (1991). Some clarifications of the CANDECOMP algorithm applied to INDSCAL. *Psychometrika*, 56, 317–326.

Ten Berge, J.M.F., & Kiers, H.A.L. (1996). Some uniqueness results for PARAFAC2. *Psychometrika*, 61, 123–132.

Ten Berge, J.M.F., & Kiers, H.A.L., Van der Stel, V. (1992). Simultaneous Components Analysis. *Statistica Applicata*, 4, 377–392.

Ten Berge, J.M.F., & Knol, D.L. (1984). Orthogonal rotations to maximal agreement for two or more matrices of different column orders. *Psychometrika*, 49, 49–55.

Tucker, L.R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Washington D.C.: Dept. of the Army.

Van Schuur, H. (1984). *Structure in political beliefs. A new model for stochastic unfolding with application to European party activists.* Unpublished doctoral dissertation, University of Groningen, Groningen.

Velicer, W.F., & Jackson, D.N. (1990a). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25, 1–28.

Velicer, W.F., & Jackson, D.N. (1990b). Component analysis versus common factoranalysis: Some further observations. *Multivariate Behavioral Research*, 25, 97–114.

Velicer, W.F., Peacock, A.C., & Jackson, D.N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research*, 17, 371–388.

Widaman, K.F. (1993). Common factor analysis versus component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, 28, 263–311.

Zwick, W.R., & Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432–442.

SAMENVATTING VAN HET PROEFSCHRIFT:

"Simultane componenten- en factor-analysemethoden voor twee of meer groepen: Een vergelijkend onderzoek"

In de sociale wetenschappen komt het vaak voor dat gegevens van twee of meer groepen personen beschikbaar zijn. Hierbij kan bijvoorbeeld gedacht worden aan persoonlijkheidsvragenlijsten, afgenomen bij steekproeven uit verschillende landen of bijvoorbeeld verschillende groepen uit de samenleving. Bij de analyse van dit soort gegevens is de doelstelling vaak om de scores op een groot aantal variabelen terug te brengen tot scores op een klein aantal zinvol interpreteerbare componenten of factoren. In het voorbeeld zijn dit persoonlijkheidscomponenten of -factoren. Vaak is het wenselijk om in dergelijke groepen dezelfde componenten of factoren te gebruiken. Hiervoor zijn simultane componenten- en factor-analysemethoden ontwikkeld.

Voor een simultane analyse van scores van de twee of meer groepen zijn verschillende analysemethoden beschikbaar, welke tot nog toe niet of nauwelijks systematisch onderzocht zijn op hun vermeende kwaliteiten, en waarvan het derhalve niet zeker is of deze methoden inderdaad doen wat ze geacht worden te doen: Het vinden van een klein aantal in twee of meer groepen aanwezige identieke componenten of factoren die aan de scores van de twee of meer groepen ten grondslag liggen.

In dit onderzoek zijn vier simultane analysemethoden nader onderzocht. Deze methoden zijn bekend als SCA-W (ook wel SCA genoemd), SCA-P, SCA-S en SIFASP (in het huidige onderzoek SIFASP-ML genoemd). Met behulp van een uitgebreid simulatieonderzoek is gekeken of deze methoden in staat waren om de in de simulatie vastgelegde factoren goed terug te vinden. In het onderzoek is steeds gezocht naar antwoord op de vraag wat een gebruiker van deze methoden het beste zou kunnen doen om een eenvoudig interpreteerbare en correcte oplossing te vinden. Dit onderzoek is derhalve vanuit het oogpunt van de consument geschreven.

Van de vier methoden is SCA-P de te preferen methode gebleken. De

methoden SCA–W en SIFASP gaven resultaten die of net zo goed of minder goed waren dan die van SCA–P en het gebruik van deze methoden wordt derhalve afgeraden. De methode SCA–S presteerde even goed als SCA–P (d.w.z. gaf doorgaans dezelfde componenten als SCA–P) wanneer de groepen uit dezelfde populatie afkomstig waren, maar presteerde beduidend minder goed dan SCA–P (gaf de 'goede' componenten vaak niet terug) wanneer de componenten uit verschillende populaties afkomstig waren. Dit resultaat maakt het mogelijk om SCA–S te gebruiken naast SCA–P om te zien of componenten uit dezelfde of uit verschillende populaties afkomstig zijn.

Dit proefschrift is als volgt opgebouwd:

In Hoofdstuk 1 wordt allereerst principale componenten–analyse voor een dataset bestaande uit één groep individuen besproken. Aan de hand hiervan worden vervolgens de simultane componenten– of factor–analyse–methoden kort geïntroduceerd. Tevens worden enkele simultane analysemethoden besproken die niet in dit onderzoek zijn opgenomen, dan wel in een vroeg stadium zijn afgevallen.

In Hoofdstuk 2 worden de vier methoden SCA–W, SCA–P, SCA–S en SIFASP–ML uitgebreid besproken. Aan elke methode is een rotatie toegevoegd om de interpretatie van de gevonden componenten of factoren te vereenvoudigen. Voor SCA–W is dit de rotatie van Hakstian, voor SCA–P en SIFASP–ML de 'Harris & Kaiser independent cluster' rotatie, en voor SCA–S een rotatie van Kiers en Ten Berge. Aan het eind van Hoofdstuk 2 wordt een eerste (voornamelijk theoretische) vergelijking tussen de methoden gegeven.

In Hoofdstuk 3 wordt de constructie van data, zoals gebruikt in het simulatieonderzoek, besproken. De variabelen worden geconstrueerd als lineaire combinaties van factoren. De gewichten voor de lineaire combinaties zijn vastgelegd in een patroonmatrix. Voor elke steekproef met gesimuleerde scores van een groep individuen op een aantal variabelen zijn het patroon, de structuur (dit is een matrix met correlaties tussen de variabelen en de factoren) en de correlaties tussen de factoren gedefinieerd (in de constructie van de data wordt enkel de term factoren gebruikt) zoals deze aanwezig zijn in de populatie waaruit de steekproef getrokken is. De geconstrueerde datasets bestaan uit twee of meer steekproeven die zijn in te delen in twee categoriën: Afkomstig uit dezelfde populatie of afkomstig uit verschillende populaties. Deze

definitie wordt besproken en er wordt een maat beschreven die de grootte van het verschil tussen populaties weergeeft.

In Hoofdstuk 4 worden de in dit onderzoek gemanipuleerde onafhankelijke variabelen besproken en worden de gekozen waarden voor de onafhankelijke variabelen opgesomd. De gemanipuleerde variabelen zijn steekproefgrootte, hoeveelheid ruis toegevoegd aan de scores op de variabelen, aantal variabelen, aantal factoren in de populatie, sterkte van de factoren, en correlaties tussen de factoren. De factoren worden in alle experimenten, met uitzondering van een experiment in Hoofdstuk 8, zo gekozen dat er in het patroon geen sprake is van overlap tussen de factoren.

Vervolgens worden in Hoofdstuk 4 maten besproken die in dit onderzoek gebruikt zijn om een indicatie te krijgen van de dimensie (het aantal factoren) die aan de scores op het (grote) aantal variabelen ten grondslag ligt. Deze dimensie is in de praktijk (bij de gebruiker) tenslotte meestal niet bekend. Voor de SCA-methoden zijn vier dimensie-indicatoren gebruikt, welke allen generalisaties zijn van maten uit de situatie waar slechts één groep geanalyseerd wordt. De vier dimensie-indicatoren zijn achtereenvolgens generalisaties van: De 'eigenwaarde groter dan 1'-regel, parallelle analyse, en de scree-test (in twee varianten). Voor SIFASP-ML zijn drie fitindices gebruikt als dimensie-indicator, te weten de 'sequential chi-square difference test', de 'root mean square error of approximation index', en de 'expected crossvalidation index'. Als vierde dimensie-indicator is een combinatie van deze drie fitindices gebruikt.

In Hoofdstuk 4 worden ook de succescriteria besproken, op basis waarvan de analysemethoden beoordeeld zijn. Ten eerste wordt de 'recovery rate' besproken. Hiermee is gekeken of de door de methoden gevonden componenten of factoren de gewenste interpretatie krijgen (d.w.z. dezelfde interpretatie als de onderliggende factoren). Ten tweede wordt gekeken of de correlaties tussen de factoren in de populatie goed teruggevonden worden door de methoden.

Aan het eind van Hoofdstuk 4 worden twee pilot-onderzoeken beschreven, welke nodig waren om zinvolle waarden voor de in het simulatieproces toegevoegde hoeveelheid ruis te bepalen, en om optimale convergentiecriteria voor de door SCA-W en SCA-S gebruikte iteratieve

procedures te bepalen.

Hoofdstuk 5 tot en met 8 beslaan de experimenten, uitgevoerd voor de onderlinge vergelijking van de vier analysemethoden. In Hoofdstuk 5 worden de drie SCA–methoden onderling vergeleken in een totaal van 576 verschillende condities. De gemanipuleerde variabelen 'steekproef–grootte', 'aantal variabelen' en 'aantal factoren' hebben in dit experiment de volgende waarden: De steekproefgrootte bedraagt 50, 100 of 150; het aantal variabelen is 12 of 24; en het aantal factoren is 2 of 4. Uit dit basisexperiment komt naar voren dat de methode SCA–P de beste resultaten geeft. Tevens wordt een discriminantanalyse gedaan om datasets met steekproeven uit dezelfde populatie te kunnen onderscheiden van datasets met steekproeven uit verschillende populaties. De gevonden discriminantfunctie blijkt (binnen de voor de discriminantanalyse gebruikte waarden van de in de discriminantfunctie opgenomen onafhankelijke variabelen) goed te werken.

In Hoofdstuk 6 wordt de methode SIFASP–ML toegepast op een selectie van 96 verschillende condities uit Hoofdstuk 5 (dezelfde datasets als in Hoofdstuk 5 worden derhalve geanalyseerd door SIFASP–ML). SIFASP–ML wordt zowel toegepast op covarianties als op correlaties. De resultaten voor covarianties komen dichter bij de onderliggende waarden dan de resultaten voor correlaties. De analyse van covarianties valt derhalve te prefereren boven de analyse van correlaties. In de vergelijking van de resultaten van SIFASP–ML met covarianties met de resultaten van SCA–P blijkt dat de laatste de beste resultaten oplevert.

In Hoofdstuk 7 worden de drie SCA–methoden getoetst op het terugvinden van de onderlinge sterkteverhouding tussen factoren, zoals deze bestaat in de populatie. Hiervoor wordt in Hoofdstuk 7 eerst een aparte maat geïntroduceerd. SCA–S blijkt de methode te zijn welke (gemiddeld over alle condities in dit experiment) het beste de onderlinge sterkte tussen factoren terugvindt.

In Hoofdstuk 8 worden de vier analysemethoden toegepast op enkele specifieke condities, te weten een conditie waarin vier groepen simultaan geanalyseerd worden, condities met steekproefgroottes van 300, 500 en 1000, en een conditie waarin de factoren in de populatie overlappen. De resultaten voor deze experimenten brengen geen verandering in de reeds getrokken conclusies.

In Hoofdstuk 9 worden de resultaten van dit onderzoek samengevat en gepresenteerd in een consumententabel, waar uit af te lezen is hoe de vier methoden relatief gesproken presteerden in de verschillende condities in dit onderzoek. Van elke methode worden apart sterke en zwakke kanten opgesomd. Vervolgens worden de beperkingen van het huidige onderzoek besproken. Hoofdstuk 9 sluit af met richtlijnen voor het gebruik van de simultane componenten–analysemethoden SCA–P en SCA–S in de praktijk. Als voorbeeld zijn in Appendix D twee empirische datasets geanalyseerd volgens deze richtlijnen.

AUTHOR INDEX

SIMULTANEOUS COMPONENT AND FACTOR ANALYSIS METHODS FOR TWO OR MORE GROUPS: A COMPARATIVE STUDY

In the social sciences one often is presented with data form two or more groups of individuals. One can think of personality inventories, administered to samples of individuals from different countries or, for instance, to different groups in society. In the analysis of these kinds of data it is often the goal to summarize the scores on a large number of variables in a small number of meaningful components or factors. It is also preferable to use the same components or factors in all groups. For this reason simultaneous component and factor analysis methods have been developed.

For a simultaneous analysis of scores of two or more groups several methods of analysis are available. Up to now, there has been almost no systematic research into the qualities of these methods. For this reason it is not certain that these methods indeed do what they are supposed to do: Finding a small number of components or factors, that are identical in the two or more groups and that adequately summarize the scores of the two or more groups.

In this book four simultaneous analysis methods are compared. These methods are known as SCA-W (also called SCA), SCA-P, SCA-S and SIFASP. In an extensive simulation study it is investigated whether these methods are capable to recover the factors underlying the data constructed by simulation.

Of the four methods, SCA-P is the method to be preferred. The methods SCA-W and SIFASP gave results that were at best equally accurate to those of SCA-P and the use of these two methods is therefore not recommended. The method SCA-S performed about the same as SCA-P (i.e. found the same components as SCA-P) when the groups came from the same population, but performed considerably less well than SCA-P (did often not recover the 'correct' components) when the components came from different populations. This result makes it possible to use SCA-S alongside SCA-P to inspect whether components are coming from equal or different populations.

Guidelines are given for the use of the simultaneous component analysis methods SCA-P and SCA-S in practice. As an exemple, two empiral data sets are analyzed following these guidelines.

DSWO PRESS

ISBN 90-6695-134-6