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.