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

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 ('$\mathbf{P}_e\mathbf{P}_e$', '$\mathbf{I}$ $\mathbf{I}$', 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.