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 $\mathbf{\Phi}$ 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 '$\mathbf{I}$ $\mathbf{I}$' 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 '$\mathbf{I}$ $\mathbf{I}$' and '$\mathbf{I}$ $\mathbf{\Phi}_{.4}$', 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*

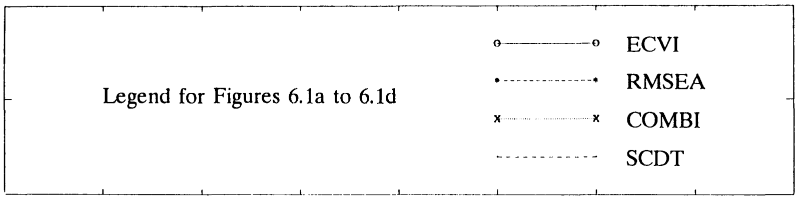
Number of factors		1	2	3	4	5	6
Proper solutions:	2 fact	458	467	129	27	0	0
	4 fact	426	437	402	367	53	5
Nonconverged sol.	2 fact	22	**11***	280	419	466	480
	4 fact	54	43	71	**105***	369	453
Nonadmissible sol.	2 fact	0	**2***	71	34	14	0
	4 fact	0	0	7	**8***	58	22

******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



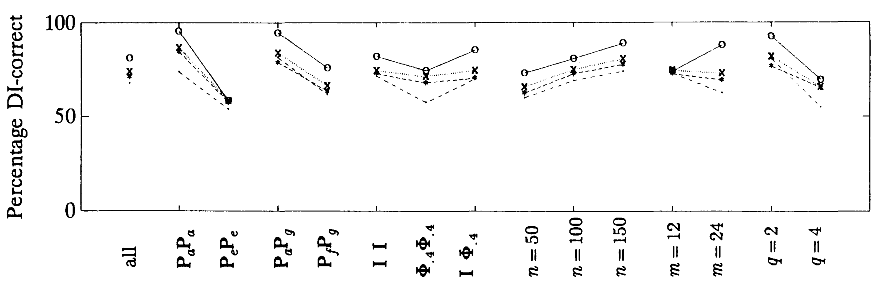


Figure 6.1a *Percentages of correct dimension indications for the method SIFASP-ML-C, averaged over each level of each independent variable, using four different dimension indicators, leaving out all nonconverged and all nonadmissible solutions*

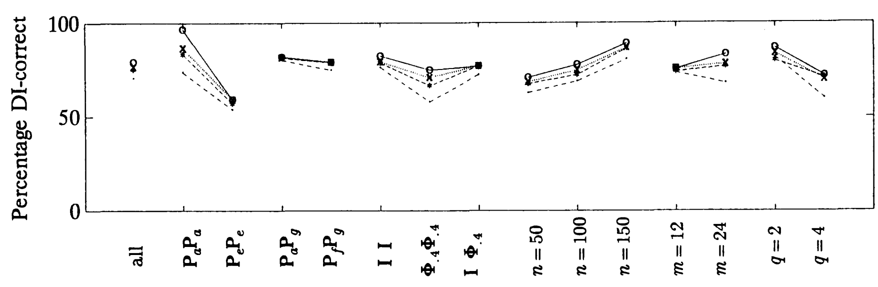


Figure 6.1b *Percentages of correct dimension indications for the method SIFASP-ML-R, averaged over each level of each independent variable, using four different dimension indicators, leaving out all nonconverged and all nonadmissible solutions*

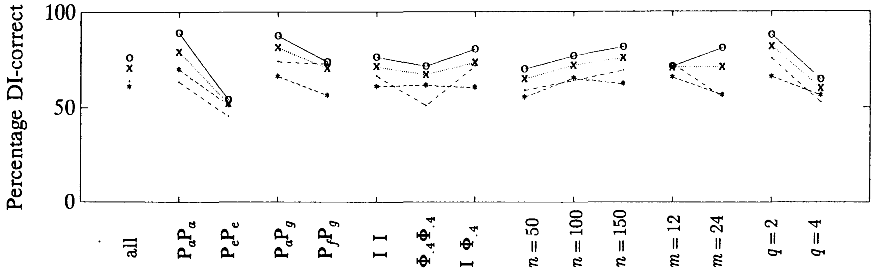


Figure 6.1c *Percentages of correct dimension indications for the method SIFASP-ML-C, averaged over each level of each independent variable, using four different dimension indicators, leaving out all nonconverged solutions*

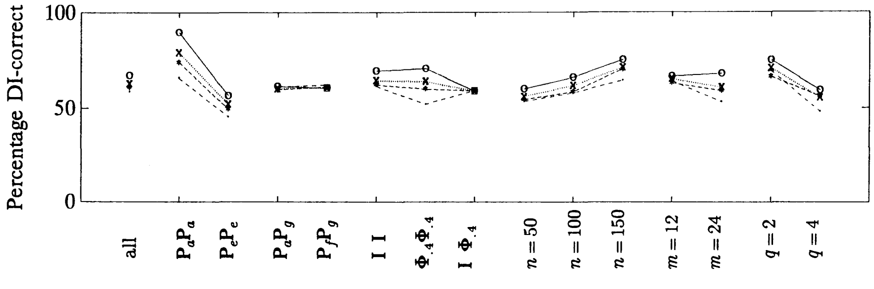


Figure 6.1d *Percentages of correct dimension indications for the method SIFASP-ML-R, averaged over each level of each independent variable, using four different dimension indicators, leaving out all nonconverged solutions*

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.

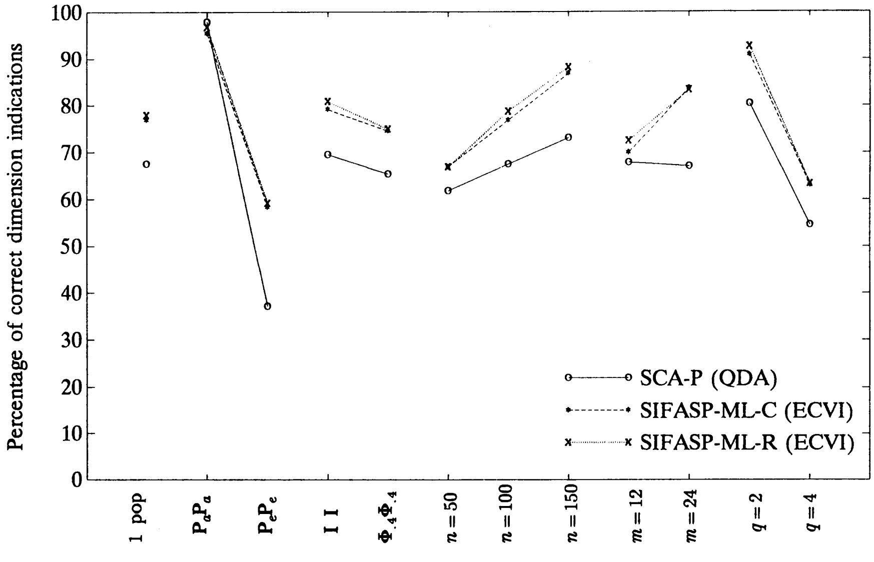


Figure 6.2a *Percentages of correct dimension indications for the two SIFASP – ML – methods (using the dimension indicator ECVI, and leaving out all nonconverged and all nonadmissible solutions) and the method SCA – P (using the dimension indicator QDA), averaged over each level of each independent variable,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

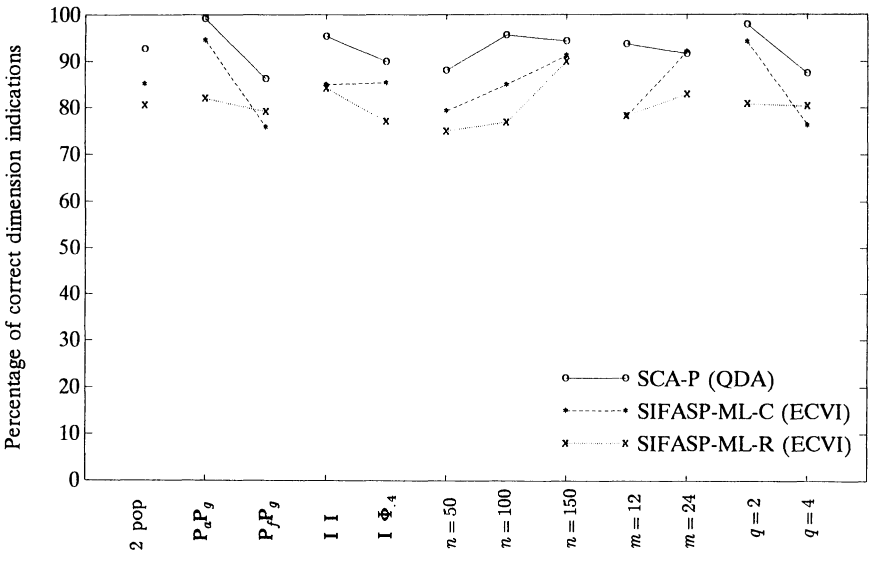


Figure 6.2b *Percentages of correct dimension indications for the two SIFASP – ML – methods (using the dimension indicator ECVI, and leaving out all nonconverged and all nonadmissible solutions) and the method SCA – P (using the dimension indicator QDA), averaged over each level of each independent variable,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

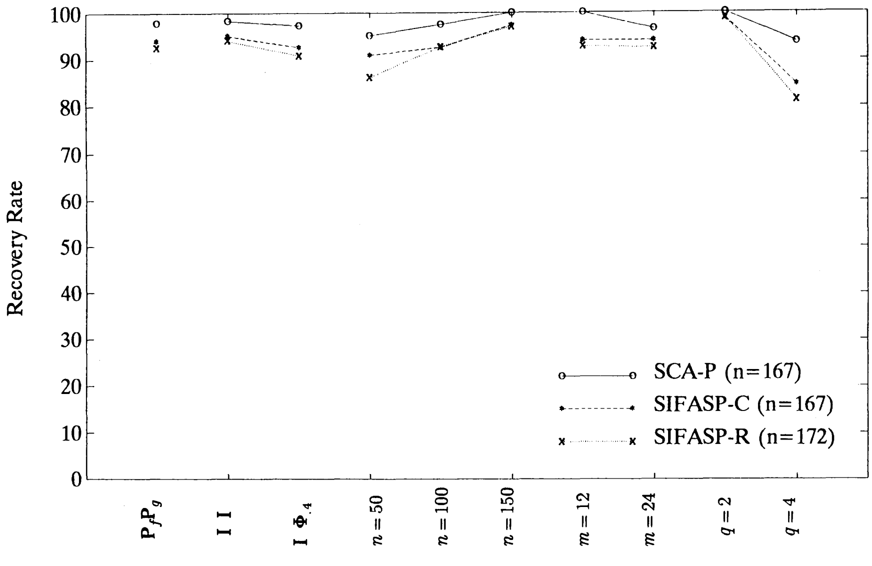


Figure 6.3a *Recovery Rates for the two SIFASP-ML-methods (leaving out all nonconverged and all nonadmissible solutions) and the method SCA-P (for the same selection of data sets as used for the method SIFASP-ML-C), averaged over each level of each independent variable, within PatternComb condition*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

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 '$\mathbf{P}_e\mathbf{P}_e$', 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

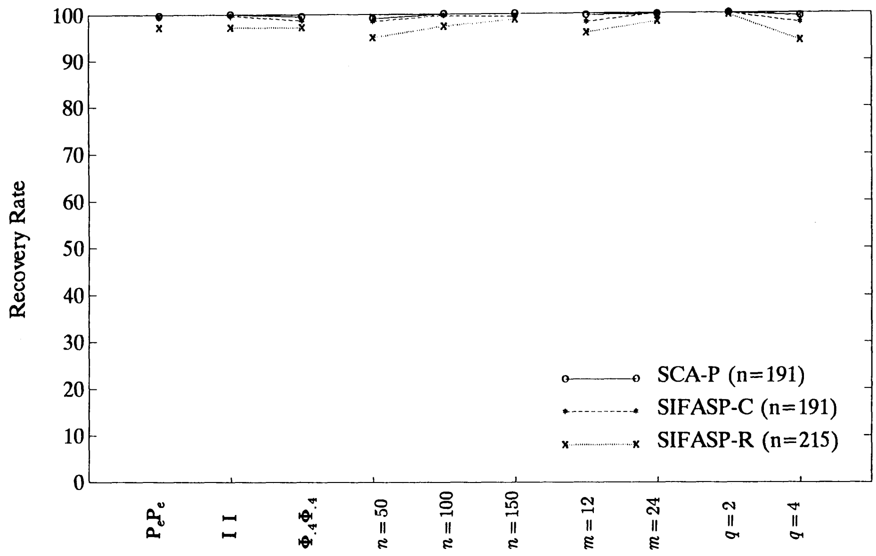


Figure 6.3b *Recovery Rates for the two SIFASP−ML−methods (leaving out all nonconverged and all nonadmissible solutions) and the method SCA−P (for the same selection of data sets as used for the method SIFASP−ML−C), averaged over each level of each independent variable, within PatternComb condition*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

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 $\mathbf{I}$ $\mathbf{I}$	.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 $\mathbf{I}$ $\mathbf{I}$	.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 $\mathbf{I}$ $\mathbf{I}$	.114 (120)	.106 (119)	.095 (119)
$\mathbf{P}_a\mathbf{P}_g$ and $\mathbf{I}$ $\mathbf{\Phi}_{.4}$	.100 (118)	.097 (119)	.105 (119)
$\mathbf{P}_f\mathbf{P}_g$ and $\mathbf{I}$ $\mathbf{I}$	.128 (109)	.127 (94)	.095 (94)
$\mathbf{P}_f\mathbf{P}_g$ and $\mathbf{I}$ $\mathbf{\Phi}_{.4}$	.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.