Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size

J. C.F. de Winter & D. Dodou

To cite this article: J. C.F. de Winter & D. Dodou (2012) Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size, Journal of Applied Statistics, 39:4, 695-710, DOI: 10.1080/02664763.2011.610445

To link to this article: https://doi.org/10.1080/02664763.2011.610445

Published online: 25 Aug 2011.

Journal of Applied Statistics
Vol. 39, No. 4, April 2012, 695–710

Taylor & Francis
Taylor & Francis Group

Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size

J.C.F. de Winter* and D. Dodou

Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

(*Received 19 February 2011; final version received 1 August 2011*)

Principal axis factoring (PAF) and maximum likelihood factor analysis (MLFA) are two of the most popular estimation methods in exploratory factor analysis. It is known that PAF is better able to recover weak factors and that the maximum likelihood estimator is asymptotically efficient. However, there is almost no evidence regarding which method should be preferred for different types of factor patterns and sample sizes. Simulations were conducted to investigate factor recovery by PAF and MLFA for distortions of ideal simple structure and sample sizes between 25 and 5000. Results showed that PAF is preferred for population solutions with few indicators per factor and for overextraction. MLFA outperformed PAF in cases of unequal loadings within factors and for underextraction. It was further shown that PAF and MLFA do not always converge with increasing sample size. The simulation findings were confirmed by an empirical study as well as by a classic plasmode, Thurstone's box problem. The present results are of practical value for factor analysts.

Keywords: exploratory factor analysis; principal axis factoring; maximum likelihood factor analysis; parameter estimation; simulations; empirical data; plasmode

Exploratory factor analysis is one of the most widely used statistical methods in psychological research [13,17]. When conducting an exploratory factor analysis, a researcher has to make decisions regarding the estimation method, the number of factors to retain, the rotation method, and the method for calculating scores (e.g. [16,17,19,42]). This article focuses on the consequences of choosing between iterative principal axis factoring (PAF) and maximum likelihood factor analysis (MLFA), two of the most commonly used factor analysis procedures [6,12].

PAF and MLFA are based on different assumptions and therefore yield different estimates of factor loadings. The $p \times p$ population correlation matrix of an orthogonal factor analysis model

*Corresponding author. Email: j.c.f.dewinter@tudelft.nl

ISSN 0266-4763 print/ISSN 1360-0532 online

http://dx.doi.org/10.1080/02664763.2011.610445
http://www.tandfonline.com

can be written as follows:

$$\Sigma = \Lambda\Lambda' + \Theta^2, \tag{1}$$

where Λ is a $p \times r$ matrix with common factor loadings and Θ is a $p \times p$ diagonal matrix with unique factor loadings. A factor analysis procedure obtains estimates of these loadings by minimizing a discrepancy function between an observed sample correlation matrix S and the model-implied (fitted) correlation matrix Σ.

PAF is a least-squares estimation of the common factor model. PAF makes no assumption about the type of error and minimizes the unweighted sum of the squares (unweighted least squares (ULS) or ordinary least squares (OLS)) of the residual matrix (e.g. [29]):

$$F_{\mathrm{OLS}} = \frac{1}{2}\mathrm{tr}[(S - \Sigma)^2] = \sum_i \sum_j (s_{ij} - \sigma_{ij})^2, \tag{2}$$

where s_{ij} and σ_{ij} are elements of the observed sample correlation matrix and the implied correlation matrix, respectively. The maximum likelihood (ML) estimation is derived from the normal distribution theory and assumes that all error is sampling error. The ML value is obtained when the following discrepancy function is minimized:

$$F_{\mathrm{ML}} = \ln|\Sigma| - \ln|S| + \mathrm{tr}[S\Sigma^{-1}] - p, \tag{3}$$

where p is the number of variables. It has been demonstrated before [33] that the ML discrepancy function is approximated by

$$F_{\mathrm{ML}} \approx \sum_i \sum_j \left[\frac{(s_{ij} - \sigma_{ij})^2}{u_i^2 u_j^2} \right], \tag{4}$$

where u_i and u_j are sample unique variances of variables i and j. As Equation (4) describes, the squared residuals are weighted by the product of the inverse of the unique variances. This is consistent with normal distribution theory, according to which weak correlations are affected by sampling error more than strong correlations and therefore the former should receive less weight.

Because ML assigns less weight to the weaker correlations, it can be expected that MLFA is less able than PAF to recover the weaker factors. This hypothesis was confirmed by Briggs and MacCallum [5] and by MacCallum *et al.* [33], who systematically investigated the recovery of weak factors when model error and sampling error were introduced either independently or in combination. In no case was there a clear advantage of MLFA over PAF, regardless of the amount and type of error introduced. Ximénez [55] also found that in a confirmatory factor analysis framework that OLS was superior in recovering weak factors, even when the model did not exactly hold in the population.

ML exhibits theoretic advantages because it is an asymptotically efficient estimator. The Cramér–Rao theorem predicts that when the variables are multinormal and the common factor model holds in the population (i.e. all error is sampling error), MLFA generates the solution that most accurately reflects the underlying population pattern. Hence, there should be circumstances under which MLFA is the preferred estimation method. Browne [7] reported a slight advantage of MLFA over PAF for recovering approximately equally strong factors with varying loadings within factors.

Despite the existing formal mathematical framework and the available articles and textbooks elaborating on the properties of PAF and MLFA [1,5,10,21,28,37,40,46], there is little practical evidence regarding which factor analysis method should be preferred for different types of factor patterns and sample sizes. By means of simulations, this article compared the factor recovery by PAF and MLFA for a variety of distortions of ideal simple structure. An empirical study was conducted as well, to investigate factor recovery in a more realistic context. Finally, Thurstone's box problem was used to investigate the behavior of the two methods for physical data with no measurement error.

Simulation study

Simulations were conducted to investigate the effect of 16 distortions from a baseline pattern matrix (level of all loadings $\lambda = 0.6$, number of factors $f = 3$, number of variables $p = 18$, no correlation between factors, and no model error) for $N = 50$. The level of loadings and number of factors of this baseline is representative of the use of factor analysis in psychological research [26]. $N = 50$ was considered a relatively small sample size yielding marginally acceptable congruence with the population loadings [15].

Method

The factor loadings of the baseline and 8 of the 16 distortions are given in Table 1. The mean of the nonzero factor loadings was 0.6 in all cases investigated. Based on the population loadings of each case, 2500 sample data matrices were generated using the method developed by Hong [27]. This is a method for introducing model error and factor correlations into a correlation matrix based on the more commonly known Tucker–Koopman–Linn procedure [49]. The correlation matrix of each sample was submitted to PAF and MLFA. All simulations were conducted with MATLAB (7.7.0.471, R2008b).

PAF began with squared multiple correlations in the diagonal, and the iterative procedure continued until the maximum absolute difference of communalities between two subsequent iterations was smaller than 10^{-3}. The maximum number of iterations was set at 9999. This tolerance interval was combined with a large number of iterations (9999) to ensure that all solutions converged toward their asymptotic value. Note that a tolerance interval of 10^{-3} with 25 iterations as well as an interval of 10^{-6} with 9999 iterations were also tested, but the results differed little from the adopted settings. The former combination yielded a lower number of Heywood cases because diverging solutions halted after 25 iterations; the latter combination slightly increased the number of Heywood cases while requiring longer computational time. For MLFA, the MATLAB function for ML factor analysis called *factoran* was applied. This function is based on Harman [24], Jöreskog [28], and Lawley and Maxwell [31], and yields virtually identical results to the ML factor analysis of SPSS. The same maximum number of iterations and tolerance interval were used as with PAF.

Three factors were extracted in Cases 1–13, one and two factors in Cases 14 and 15, respectively (underextraction), and four and five factors in Cases 16 and 17, respectively (overextraction). The loadings were rotated with an oblique Procrustes rotation to achieve the best transformation that fits the population pattern and to avoid rotational indeterminacy. To recover the order and sign of the loadings, Tucker's congruence coefficients (K; [48]) were calculated for the nine loading vector combinations between the sample loading matrix and the population loading matrix. Next, the reordering of the sample loadings started with the highest absolute K and proceeded toward the lowest K until the sign and order of the factors were recovered. For under- and overextraction, in which the number of extracted factors was, respectively, lower or higher than the number of factors of the population loading matrix, the recovery of loadings was evaluated for the combination of loadings with the highest K for the minimum number of factors. That is, one or two factors for underextraction and three factors for overextraction.

If a sample yielded a variable with communality greater than 0.998, then this solution was labeled a Heywood case. If a Heywood case occurred, the sample loading matrices of both PAF and MLFA were discarded. Sample loading matrices of both methods were also discarded in case that the rotation did not converge.

Table 1. Population loading matrices used in the simulations.

Case 1			Case 4			Case 5			Case 7		
Baseline			Zero loadings replacing nonzero loadings			Reduced number of variables			Unequal loadings between factors		
0.6	0	0	0.6	0	0	0.6	0	0	0.9	0	0
0.6	0	0	0.6	0	0	0.6	0	0	0.9	0	0
0.6	0	0	0.6	0	0	0.6	0	0	0.9	0	0
0.6	0	0	0.6	0	0	0	0.6	0	0.9	0	0
0.6	0	0	0	0.6	0	0	0.6	0	0.9	0	0
0.6	0	0	0	0.6	0	0	0.6	0	0.9	0	0
0	0.6	0	0	0.6	0	0	0	0.6	0	0.6	0
0	0.6	0	0	0.6	0	0	0	0.6	0	0.6	0
0	0.6	0	0	0	0.6	0	0	0.6	0	0.6	0
0	0.6	0	0	0	0.6				0	0.6	0
0	0.6	0	0	0	0.6				0	0.6	0
0	0.6	0	0	0	0.6				0	0.6	0
0	0	0.6	0	0	0				0	0	0.3
0	0	0.6	0	0	0				0	0	0.3
0	0	0.6	0	0	0				0	0	0.3
0	0	0.6	0	0	0				0	0	0.3
0	0	0.6	0	0	0				0	0	0.3
0	0	0.6	0	0	0				0	0	0.3

Cases 8 and 9			Case 10			Case 11			Case 13		
Unequal loadings within factors			Crossloadings			Unequal p/f (moderate)			Random variation of all loadings		
0.9	0	0	0.6	0.6	0	0.6	0	0	0.726	0.117	0.078
0.3	0	0	0.6	0.6	0	0.6	0	0	0.762	0.184	−0.073
0.9	0	0	0.6	0.6	0	0.6	0	0	0.451	0.062	0.180
0.3	0	0	0.6	0.6	0	0.6	0	0	0.765	−0.186	−0.186
0.9	0	0	0.6	0	0	0.6	0	0	0.653	0.140	−0.025
0.3	0	0	0.6	0	0	0.6	0	0	0.439	0.174	−0.047
0	0.9	0	0	0.6	0	0	0.6	0	−0.089	0.672	0.106
0	0.3	0	0	0.6	0	0	0.6	0	0.019	0.703	0.118
0	0.9	0	0	0.6	0	0	0.6	0	0.183	0.697	−0.125
0	0.3	0	0	0.6	0	0	0.6	0	0.186	0.557	−0.004
0	0.9	0	0	0.6	0	0	0.6	0	−0.137	0.662	−0.022
0	0.3	0	0	0.6	0	0	0.6	0	0.188	0.469	0.059
0	0	0.9	0	0	0.6	0	0	0.6	0.183	0.082	0.684
0	0	0.3	0	0	0.6	0	0	0.6	−0.006	−0.187	0.702
0	0	0.9	0	0	0.6	0	0	0.6	0.120	−0.089	0.510
0	0	0.3	0	0	0.6				−0.143	−0.182	0.672
0	0	0.9	0	0	0.6				−0.031	−0.161	0.662
0	0	0.3	0	0	0.6				0.166	0.129	0.465

Note: For Case 13, a different population pattern was produced for each repetition.

Independent variables. Factor recovery by PAF and MLFA was studied for the following 17 cases:

Case 1: Baseline ($\lambda = 0.6$, $f = 3$, $p = 18$).

Case 2: Correlated factors. Many psychological constructs are correlated with each other. We investigated factor recovery when all three combinations of factors were substantially (0.7) correlated.

Case 3: Model error. Model error represents lack of fit of the factor model to the population due to phenomena such as the presence of a large number of minor factors and the nonlinear influence of common factors on measured variables. Random model error was introduced for every sample by means of 200 minor factors, explaining 20% of the variance. The parameter determining the distribution of the minor factors was set at 0.25, which implies that the minor factors were considerably skewed in favor of the earlier factors in the sequence.

Case 4: Zero loadings replacing nonzero loadings. Six of the 18 nonzero loadings were set to zero.

Case 5: Reduced number of variables. The number of variables affects factor recovery, e.g. [15]. Conway and Huffcutt [12] examined factor analytic practices in a variety of journals and showed that more than half of the studies used between 3 and 6 variables per factor. For our three-factor pattern we reduced the number of variables from 18 to 9.

Case 6: Increased number of variables. The number of variables was increased from 18 to 36.

Case 7: Unequal loadings between factors. The nonzero loadings of the three factors were changed to 0.9, 0.6, and 0.3, respectively.

Case 8: Unequal loadings within factors. The nonzero loadings within each factor were alternated between 0.9 and 0.3.

Case 9: Unequal loadings within correlated factors. This was the same as Case 8, but aggravated by correlating all three combinations of factors with 0.7.

Case 10: Crossloadings. Crossloadings (also called secondary loadings) are inevitable in real data and it is common practice to discard items on this ground during scale construction. Here, we compared MLFA with PAF in the presence of four strong (0.6) crossloadings.

Case 11: Unequal p/f (moderate). In reality an equal p/f occurs only rarely. Therefore, the third factor was weakened by decreasing the number of variables per factor from 6 to 3.

Case 12: Unequal p/f (severe). The third factor was weakened by decreasing the number of variables per factor from 6 to 2.

Case 13: Random variation of all loadings. In empirical data, population loadings are not homogeneous. For each sample, a uniform random variation (range -0.2 to 0.2) was added to the baseline of population loadings.

Case 14: Underextraction (-2 factors). One factor was extracted instead of three. Estimating the correct number of factors is a challenge. There are a variety of methods that can assist in this decision, such as Velicer's Minimum Average Partial [38], Bayesian Information Criterion [25], parallel analysis [38], and generalized degrees of freedom [10]. Yet, no method is infallible. It is thus important to compare estimation methods when an inappropriate number of factors have been extracted.

Case 15: Underextraction (-1 factor). Two factors were extracted instead of three.

Case 16: Overextraction ($+1$ factor). Four factors were extracted instead of three.

Case 17: Overextraction ($+2$ factors). Five factors were extracted instead of three.

After the simulations with $N = 50$, factor recovery was investigated for eight different sample sizes between 25 and 5000 (2500 sample data matrices generated per sample size) for the cases which showed large differences between PAF and MLFA on the dependent variables.

Dependent variables. Differences between PAF and MLFA were assessed by means of three dependent variables:

1. Cohen's d (i.e. the standardized mean difference) between the g of PAF and the g of MLFA, with g defined as the root mean-squared difference between the sample and population loading matrices [52]:
$$g = \sqrt{\frac{\text{trace}\left[(\Lambda - \hat{\Lambda})'(\Lambda - \hat{\Lambda})\right]}{pr}}, \tag{5}$$
where $\hat{\Lambda}$ is the estimate of the loadings. A negative d indicates that MLFA performed better than PAF.
2. The proportion of solutions for which the g of PAF was greater than the g of MLFA. A proportion greater than 0.5 indicates that MLFA performed better than PAF. These first two dependent variables are complementary: d measures the size of the difference between the two methods but is sensitive to potential outliers. The proportion of outperforming solutions, in contrast, indicates which method performed better without providing information about the size of the difference, and it is unaffected by the distribution of g.
3. The proportion of improper solutions (Heywood cases) generated by PAF and MLFA.

Results

Table 2 shows the results of the simulations for $N = 50$. Heywood cases occurred most frequently for correlated factors (Case 2), a low number of variables (Case 5), unequal loadings between factors (Case 7), unequal number of variables per factor (Cases 11 and 12), and overextraction (Cases 16 and 17). Overall, MLFA was more prone to Heywood cases than PAF.

The 16 distortions had distinct effects on factor recovery, with g values above 0.20 observed for underextraction of two factors (Case 14) and correlated factors (Cases 2 and 9). For a larger number of variables, unequal loadings within factors, crossloadings, and random variation of all loadings (Cases 6, 8, 10, and 13) both methods exhibited better factor recovery performance compared with the baseline.

For simple patterns, that is, orthogonal factors, equal salient loadings on all factors, and one salient loading per factor (Cases 1, 3, 5, and 6), PAF yielded more accurate estimates of the population pattern than MLFA. Also, when one or more factors were relatively weak (Cases 4, 7, 11, and 12), PAF was better able to recover the weak factors than MLFA, which is consistent with the earlier research on the same topic. An illustration of the factor recovery per loading coefficient of Case 11 (moderately unequal p/f) is provided in Table 3. It can be seen that MLFA had poorer recovery of the third factor, which had only three salient loadings. PAF was also better than MLFA in the case of random variation of loadings (Case 13).

MLFA outperformed for some of the severe distortions, namely, unequal loadings (Case 8), especially when these were within correlated factors (Case 9), and underextraction (Cases 14 and 15). An illustration of Case 9 is provided in Table 3, showing that the loading coefficients of PAF were inaccurate, particularly for the lower loadings of the population solution. Figure 1 also illustrates the dramatic advantage of MLFA over PAF in Case 9. MLFA did not outperform PAF in the instance of correlated factors (Case 2) and crossloadings (Case 10).

Figure 2 shows the results for Cases 9 (unequal loadings within correlated factors), 11 (unequal p/f, moderate), 12 (unequal p/f, severe), 14 (underextraction, -2 factors), and 17 (overextraction,

Table 2. Proportions of Heywood cases (HW), average root mean-squared error of the factor loadings (g), proportion of MLFA outperforming solutions, and Cohen's d between the g values of PAF and MLFA, for $N = 50$.

	Case	HW PAF	HW MLFA	g PAF	g MLFA	g PAF > MLFA	d
1.	Baseline	0.00	0.04	0.124	0.129	0.134	0.27
2.	Correlated factors	0.03	0.48	0.210	0.216	0.250	0.12
3.	Model error	0.00	0.07	0.167	0.177	0.156	0.22
4.	Zero loadings replacing nonzero loadings	0.02	0.38	0.159	0.161	0.452	0.06
5.	Reduced number of variables	0.25	0.59	0.133	0.142	0.138	0.32
6.	Increased number of variables	0.00	0.00	0.117	0.118	0.150	0.13
7.	Unequal loadings between factors	0.06	0.44	0.144	0.152	0.229	0.23
8.	Unequal loadings within factors	0.01	0.02	0.109	0.106	0.742	−0.22
9.	Unequal loadings within correlated factors	0.03	0.17	0.216	0.161	0.954	−1.10
10.	Crossloadings	0.00	0.06	0.119	0.123	0.219	0.25
11.	Unequal p/f (moderate)	0.04	0.35	0.132	0.139	0.146	0.31
12.	Unequal p/f (severe)	0.17	0.59	0.148	0.162	0.167	0.36
13.	Random variation of all loadings	0.00	0.06	0.118	0.120	0.367	0.11
14.	Underextraction (−2 factors)	0.00	0.00	0.242	0.187	0.957	−0.93
15.	Underextraction (−1 factor)	0.00	0.01	0.187	0.160	0.927	−0.73
16.	Overextraction (+1 factor)	0.09	0.60	0.144	0.150	0.227	0.33
17.	Overextraction (+2 factors)	0.25	0.88	0.165	0.172	0.307	0.32

+2 factors) for sample sizes between 25 and 5000. In the presence of a weak factor (Case 11), PAF and MLFA converged with increasing sample size. When the inequality of factors was severe (Case 12) or too many factors were extracted (Case 17), however, increasing the sample size benefited PAF more than MLFA. For unequal loadings within correlated factors (Case 9), PAF and MLFA converged with increasing N, with MLFA having a slight advantage, whereas for the severe model misspecification expressed by underextracting two factors (Case 14), the methods diverged, with MLFA exhibiting a pronounced advantage over PAF for large sample sizes. In fact, in the case of underextraction, almost 100% of the MLFA solutions were better than the PAF solutions, for all investigated sample sizes.

In all simulated conditions, Heywood cases were more frequent for MLFA than for PAF and decreased with sample size. The factor recovery performance of MLFA was associated with the occurrence of Heywood cases: For underextraction, in which MLFA performed particularly well, almost no Heywood cases occurred, whereas for highly unequal p/f and overextraction, in which MLFA underperformed, Heywood cases occurred for MLFA even for large sample sizes.

Empirical study

The simulated data investigated in the previous section were well conditioned and did not include combinations of distortions and non-normally distributed variables. To investigate factor recovery under more realistic conditions, an empirical study was conducted.

Method

The data set consisted of 9098 participants who, in the framework of a large British study investigating the training, testing, and subsequent experiences of newly licensed drivers [47,53],

Table 3. Average root mean-squared error (g) per loading coefficient for Case 11 (moderately unequal p/f) and 9 (unequal loadings within correlated factors) for PAF and MLFA.

PAF			MLFA		
Case 11					
0.116	0.123	0.154	0.124	0.125	0.162
0.118	0.126	0.154	0.126	0.127	0.163
0.120	0.125	0.154	0.131	0.127	0.167
0.119	0.127	0.156	0.128	0.129	0.166
0.120	0.124	0.160	0.128	0.126	0.170
0.116	0.127	0.162	0.126	0.128	0.170
0.129	0.121	0.152	0.132	0.131	0.164
0.128	0.121	0.151	0.129	0.130	0.162
0.127	0.122	0.153	0.129	0.133	0.167
0.124	0.118	0.152	0.126	0.127	0.160
0.127	0.120	0.153	0.129	0.130	0.163
0.125	0.124	0.152	0.126	0.133	0.163
0.119	0.114	0.144	0.125	0.120	0.177
0.121	0.117	0.148	0.126	0.121	0.181
0.121	0.123	0.148	0.128	0.127	0.184
Case 9					
0.169	0.239	0.244	0.122	0.122	0.123
0.225	0.230	0.240	0.197	0.191	0.202
0.172	0.242	0.243	0.126	0.125	0.123
0.222	0.230	0.236	0.192	0.194	0.195
0.169	0.238	0.243	0.123	0.122	0.123
0.220	0.231	0.233	0.190	0.194	0.196
0.241	0.172	0.251	0.123	0.122	0.123
0.224	0.224	0.232	0.191	0.192	0.194
0.240	0.173	0.253	0.121	0.124	0.126
0.234	0.225	0.227	0.199	0.197	0.193
0.241	0.171	0.249	0.124	0.124	0.123
0.229	0.218	0.234	0.193	0.188	0.195
0.241	0.250	0.175	0.123	0.123	0.124
0.230	0.236	0.227	0.194	0.195	0.192
0.247	0.249	0.176	0.123	0.124	0.126
0.230	0.235	0.219	0.191	0.199	0.193
0.246	0.246	0.173	0.125	0.122	0.125
0.236	0.229	0.224	0.196	0.195	0.199

Note: Gradient scale visualizes the size of the error, from small to large.

filled in the Driver Behaviour Questionnaire (DBQ). The participants responded to the query “when driving, how often do you do each of the following?” with respect to 38 items, such as “attempt to drive away from traffic lights in too high a gear” and “overtake a slow driver on the inside”. The responses were on a 6-point scale ranging from “*never*” to “*nearly all the time*”. The DBQ has a well-established factor structure distinguishing between errors and violations during driving. The DBQ is probably the most popular questionnaire used in road safety research and has been used in at least 174 studies [14].

For each factor, the six variables with the highest loadings on this factor were selected, resulting in a 9098 × 12 data matrix. The four largest eigenvalues of the unreduced 12 × 12 correlation matrix were 3.16, 2.29, 1.01, and 0.78, supporting the two-factor structure. The PAF and MLFA solutions (oblimin rotation with gamma = 0, interfactor correlation = 0.15) were nearly identical ($g = 0.009$). We used the average of these as reference matrix (see Table 4) and conducted a sampling study with eight sample sizes between 25 and 5000. For each sample size, 2500 random samples were drawn without replacement and factor analyzed as in the simulation study. The loadings were rotated with an oblique Procrustes rotation.

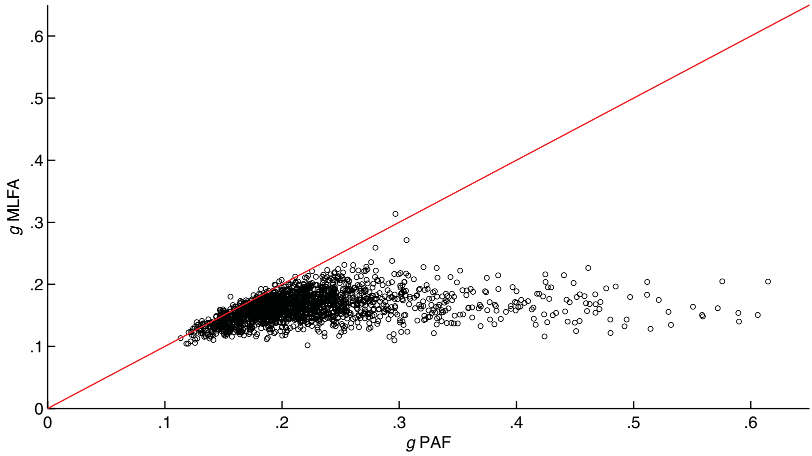

Figure 1. Results for simulation Case 9 (unequal loadings within correlated factors). Root mean-squared error (g) for MLFA versus g for PAF. Solutions yielding Heywood cases or rotations that failed to converge were omitted, resulting in 1847 of 2500g values. The line of unity is depicted for reference.

Factor recovery was also investigated when weakening the violations factor (similar to Case 12 of the simulation study) by choosing two instead of six variables with lower loadings on that factor, resulting in the reference matrix shown in Table 5. The g between PAF and MLFA loadings was 0.006 and the correlation between factors 0.08. Finally, factor recovery was investigated for underextraction, that is, when extracting one instead of two factors (similar to Cases 14 and 15 of the simulation study).

Results

The factor recovery results are shown in Figure 3. When extracting two factors, PAF was slightly superior, in agreement with the small advantage of PAF observed in the simulations. When the violations factor was weakened, PAF outperformed more clearly. MLFA was superior in underextraction, with the discrepancy between the two methods increasing with sample size. This can also be seen by comparing Tables 4 and 6: PAF dissolved the errors and violations factors, whereas the extracted MLFA factor more closely corresponded to the violations factor of the reference matrix of Table 4.

Plasmode study

Apart from behavioral sciences, factor analysis is becoming increasingly common in domains such as chemistry, climate research, and engineering, areas in which data are characterized by little measurement error [30,35,50]. In order to investigate the behavior of PAF and MLFA with such physical data, we used a plasmode. This term was coined by Cattell and Sullivan [9]; see also [8] to describe numerical data sets derived from real, preferably physical, measurements instead of simulations, and constructed so that their properties are known and that they fit an established scientific model; such data sets are then used to test the validity of a statistical technique (e.g. factor analysis) for that scientific model. One of the most well-known plasmodes is Thurstone's box problem [22], which we used herein to compare PAF with MLFA.

Method

Twenty-five virtual boxes were generated with dimensions x, y, and z uniformly random between 1 and 10. Then $2x$ was added to y so that the x and y dimensions became highly correlated (0.89

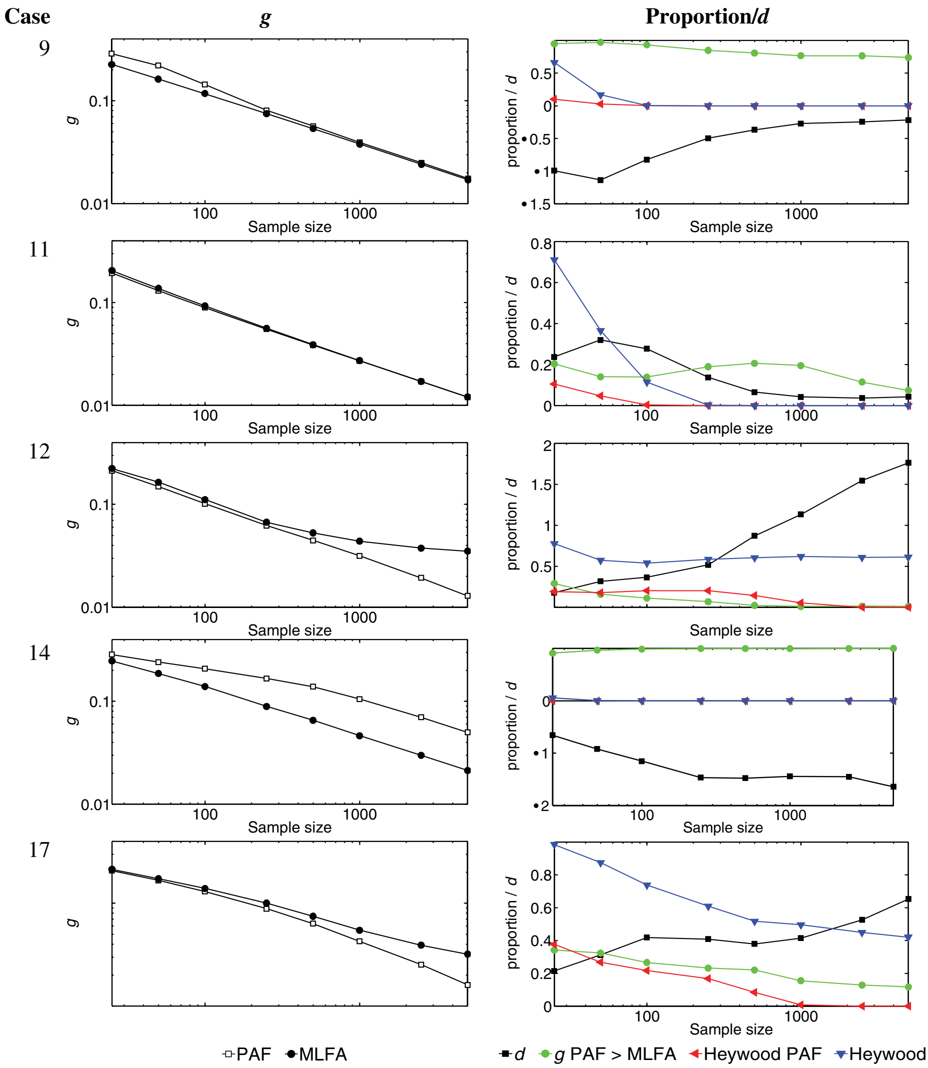

Figure 2. Results for Cases 9 (unequal loadings within correlated factors), 11 (moderately unequal p/f), 12 (severely unequal p/f), 14 (underextraction, -2 factors), and 17 (overextraction, $+2$ factors) versus sample size. Left: Root mean-squared error of loadings (g) for PAF and MLFA. Right: Cohen's d between g of PAF and g of MLFA, proportion of MLFA outperforming solutions, and proportion of Heywood cases for PAF and MLFA. A negative d and/or a proportion greater than 0.5 mean that MLFA performed better than PAF.

in the population), and 20 variables were appended such as the area and diagonals of each side, the volume of the box, and arbitrary functions of the three main dimensions [45, p. 142]: x^2, y^2, z^2, xy, xz, yz, $(x^2+y^2)^{1/2}$, $(x^2+z^2)^{1/2}$, $(y^2+z^2)^{1/2}$, $2x+2y$, $2x+2z$, $2y+2z$, $\log(x)$, $\log(y)$, $\log(z)$, xyz, $(x^2+y^2+z^2)^{1/2}$, e^x, e^y, and e^z. Thurstone's box problem contains no measurement error and therefore very high loadings are present, generating a pattern with unequal loadings within correlated factors as in Case 9 of the simulations.

The resulting 25×23 data matrix was factor analyzed as in the empirical study, and the loadings were rotated with an oblique Procrustes rotation toward the true loadings (shown in Table 7, notice the heterogeneous loadings). The procedure was repeated 2500 times. MATLAB's *factoran* could not be used as it issued a warning about the covariance matrix being positive definite. To

Table 4. Loading coefficients of the DBQ factor solution (based on 9098 × 12 data matrix).

		Violations	Errors
1.	Race away from traffic lights to beat the driver next to you	0.661	−0.051
2.	Disregard the speed limit on a motorway	0.638	0.007
3.	Disregard the speed limit on a residential road	0.633	0.137
4.	Become angered by another driver and indicate hostility	0.575	−0.026
5.	Use a mobile phone without a hands free kit	0.554	−0.017
6.	Sound your horn to indicate your annoyance	0.517	−0.082
7.	Change into the wrong gear when driving along	−0.033	0.692
8.	Drive in either too high or too low a gear for the conditions	0.076	0.660
9.	Have to confirm you are in the right gear	−0.074	0.604
10.	Select wrong gear when wanting to go into reverse	−0.063	0.493
11.	Switch on one thing when meant to switch on something else	0.009	0.467
12.	Get into the wrong lane when approaching a roundabout or junction	0.159	0.456

Note: The loading coefficients represent the average of PAF and MLFA. Gradient scale visualizes the size of the loading, from low to high.

Table 5. Loading coefficients of the DBQ factor solution (based on 9098 × 7 data matrix) in the case of weakening the violations factor.

		Violations	Errors
1.	Drive when suspect you may be over legal alcohol limit	0.381	0.012
2.	Overtake a slow driver on the inside	0.362	−0.005
3.	Change into the wrong gear when driving along	−0.025	0.711
4.	Drive in either too high or too low a gear for the conditions	0.080	0.635
5.	Have to confirm you are in the right gear	−0.025	0.599
6.	Select wrong gear when wanting to go into reverse	−0.037	0.491
7.	Switch on one thing when meant to switch on something else	−0.009	0.467

Notes: The loading coefficients represent the average of PAF and MLFA. The variable "sound your horn to indicate your annoyance" loading low only on the errors factor (Table 4) was also removed. Gradient scale visualizes the size of the loading, from low to high.

circumvent this problem, we used the MLFA MATLAB code provided by Marcus [34] see also [39]. According to Marcus, the code yields the same results as the equations reported in Morrison [36, pp. 308–309].

Results

MLFA had a clear advantage over PAF when recovering the solution of Thurstone's box problem with the correlated x and y dimension (Cohen's $d = -1.28$; proportion of outperforming MLFA solutions = 0.992). These findings are in agreement with the simulation results for Case 9 (unequal loadings within correlated factors, Cohen's $d = -1.10$; proportion of outperforming MLFA solutions = 0.954). By keeping x and y uncorrelated (i.e. by not adding $2x$ to y), MLFA preserved only a small advantage (Cohen's $d = -0.18$, proportion of outperforming MLFA solutions = 0.869).

Discussion

This study compared the performance of iterative PAF with that of MLFA. The key results are discussed below.

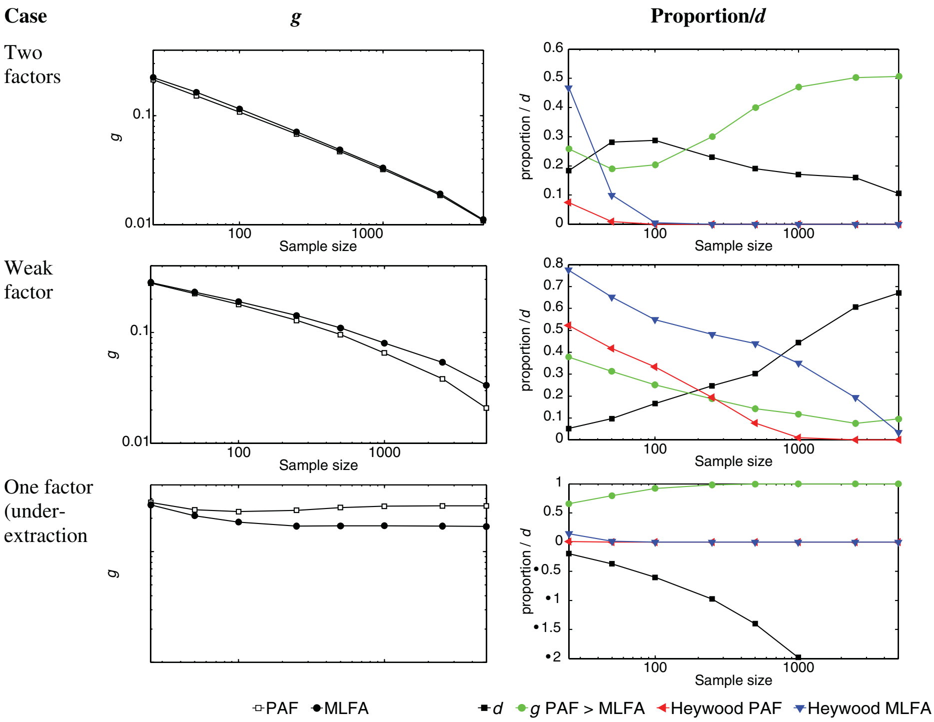

Figure 3. Results for the empirical sampling study. Left: Root mean-squared error of loadings (g) for PAF and MLFA. Right: Cohen's d between g of PAF and g of MLFA, proportion of MLFA outperforming solutions, and proportion of Heywood cases for PAF and MLFA. A negative d and/or a proportion greater than 0.5 mean that MLFA performed better than PAF. In the case of underextraction, the y-axis ranges from -2 to 1 for the sake of clarity. The minimum value of d is -5.88 for $N = 5000$.

Table 6. Loading coefficients of the DBQ factor solution (based on 9098×12 data matrix) for one extracted factor (underextraction).

		PAF	MLFA
1.	Race away from traffic lights to beat the driver next to you	0.522	0.606
2.	Disregard the speed limit on a motorway	0.546	0.631
3.	Disregard the speed limit on a residential road	0.643	0.690
4.	Become angered by another driver and indicate hostility	0.482	0.527
5.	Use a mobile phone without a hands free kit	0.468	0.531
6.	Sound your horn to indicate your annoyance	0.395	0.447
7.	Change into the wrong gear when driving along	0.384	0.244
8.	Drive in either too high or too low a gear for the conditions	0.462	0.333
9.	Have to confirm you are in the right gear	0.305	0.179
10.	Select wrong gear when wanting to go into reverse	0.258	0.147
11.	Switch on one thing when meant to switch on something else	0.309	0.204
12.	Get into the wrong lane when approaching a roundabout or junction	0.434	0.342

Note: Gradient scale visualizes the size of the loading, from low to high.

- For simple patterns (equal loadings, no crossloadings, no factor correlations), PAF is a superior estimation method. The reason is that MLFA weights each correlation differently, thereby introducing sampling variability into the minimum discrepancy function. PAF, in contrast, weights all residuals equally and therefore yields more stable loadings.

Table 7. True loading coefficients of the Thurstone's box problem for uniform random x, y, and z dimensions between 1 and 10.

	Factor 1	Factor 2	Factor 3
x	1.000	0.000	0.000
y	0.000	1.000	0.000
z	0.000	0.000	1.000
x^2	0.978	0.000	0.000
y^2	0.000	0.978	0.000
z^2	0.000	0.000	0.978
xy	0.671	0.671	0.000
xz	0.671	0.000	0.671
yz	0.000	0.671	0.671
$(x^2+y^2)^{1/2}$	0.695	0.695	0.000
$(x^2+z^2)^{1/2}$	0.695	0.000	0.695
$(y^2+z^2)^{1/2}$	0.000	0.695	0.695
$2x+2y$	0.707	0.707	0.000
$2x+2z$	0.707	0.000	0.707
$2y+2z$	0.000	0.707	0.707
$\log(x)$	0.963	0.000	0.000
$\log(y)$	0.000	0.963	0.000
$\log(z)$	0.000	0.000	0.963
xyz	0.519	0.519	0.519
$(x^2+y^2+z^2)^{1/2}$	0.565	0.565	0.565
e^x	0.720	0.000	0.000
e^y	0.000	0.720	0.000
e^z	0.000	0.000	0.720

Note: Gradient scale visualizes the size of the loading, from low to high.

- In agreement with earlier research on the same topic (e.g. [5]), PAF outperformed MLFA in recovering factors with low loadings.
- A noteworthy finding was observed for correlated factors. For $N = 50$, factor correlations of 0.7 and equal loadings, we found a slight advantage of PAF over MLFA, similar to Beauducel [3]. However, when correlated factors (0.7) were combined with unequal loadings within factors, MLFA strongly outperformed PAF (Cohen's $d = -1.10$). This finding was confirmed by the results of Thurstone's box problem.
- No clear differences between MLFA and PAF were found for crossloadings. Indeed, because all factors were equally strong, no advantage of MLFA should be expected.
- PAF and MLFA exhibited opposite tendencies regarding over- and underfactoring: PAF was better in overextraction and MLFA in underextraction. This can be explained by the fact that the overextracted factors are essentially nonexistent in the population, negatively affecting MLFA [21]. By omitting true variance, underextraction obscures the extracted factors and misspecifies the model [11]. It should be noted here that MLFA did not perform that differently from PAF in the presence of model error (Case 3). In summary, PAF performed poorly under the severe misspecification caused by failing to retain a major factor, but performed approximately equally as well as MLFA when distorting the population by means of multiple minor factors.

We are not aware of any previous comparison between PAF and MLFA for over- and underextraction. Past studies have primarily focused on a comparison between factor analysis and principal component analysis (PCA). For example, Lawrence and Hancock [32] found that when only one factor underlay the data, overfactoring hardly affected the solutions; for solutions with more than one factor, PAF was more robust than PCA. Fava and Velicer [18] found that underextraction was more detrimental for MLFA than for PCA. Note that PCA and factor analysis have different aims (data reduction versus explaining the correlation matrix) and that PCA loadings are higher, particularly when the loadings or the number of variables per factor is low [23,43]. Differences between PCA and factor analysis also occur when loadings within a factor are unequal (i.e. heterogeneous), especially when the factors are correlated [54].

- In contrast to what might be expected, PAF and MLFA solutions did not always converge with increasing sample size. The difference between the two methods grew to the advantage of PAF in the case of poorly defined factors (such as for highly unequal p/f and overextraction), and to the advantage of MLFA for misspecified models obtained by severe underfactoring.
- MLFA, as implemented in the current study, generated much more Heywood cases than PAF. The factor recovery performance of MLFA was associated with the occurrence of Heywood cases. In the cases in which MLFA performed well, almost no Heywood cases occurred, whereas for highly unequal p/f, in which MLFA underperformed, Heywood cases occurred even for the largest sample sizes. These results extend the existing knowledge base on the occurrence of Heywood cases for different estimation methods [2,4,5,20,55]. Note that the occurrence of Heywood cases need not to be seen as a sign of poor factor recovery performance *per se*. Heywood cases may indicate that the data violate assumptions of the factor model, and therefore have a diagnostic value indicating that a factor analysis should not have been conducted in the first place [51].

One limitation of this study is that we did not compare the performance of the estimation methods in confirmatory factor analysis or structural equation modeling. MLFA is better formalized in a statistical framework and allows for the calculation of fit indices, significance testing, and the estimation of confidence intervals. A simulation study by Olsson *et al*. [41] found that maximum likelihood was more stable and less biased in terms of both empirical fit and recovery of the underlying structure as compared with generalized least squares. Although extending our findings to a confirmatory framework should be done with care, one may imply from the results by Olsson *et al*. that tendencies similar to the present findings may be expected for confirmatory factor analysis.

In conclusion, the present results show that PAF generally outperforms, except for complex cases of model misspecification and unequal loadings within factors, particularly when these factors are correlated. In other words, for a relatively simple factor pattern, or when weak factors are present, PAF is the method of preference. MLFA, in contrast, is the most flexible method and best able to cope with severe model misspecifications. We found that large differences (d larger than 0.5) existed between MLFA and PAF in favor of MLFA, an observation which runs counter to Briggs and MacCallum [5] (see introduction).

The present results are relevant for practitioners who must decide which estimation method to use in exploratory factor analysis. A useful strategy in conducting exploratory factor analysis is to apply multiple estimation methods and investigate whether there are substantial differences [44]. If this is the case and the researcher is interested in recovering all relevant factors, PAF must be preferred. However, if the factor solution reveals correlated factors combined with highly unequal loadings within a factor – a situation that may be more likely in physical than behavioral data – then the MLFA solution will probably describe the population pattern more accurately. Lastly, a final note on underextraction: what should a factor analyst do if there is an indication (e.g. from the Scree plot) that a major factor is not retained? Although it is legitimate to retain a larger number of factors and use PAF, such a strategy is not always effective, as the "correct" number of factors can never be defined indisputably. Besides, a researcher may not wish to extract more factors, for example, because of theoretical considerations supporting the chosen number of factors. In such a case, MLFA must be preferred.

References

[1] T.W. Anderson and Y. Amemiya, *The asymptotic normal distribution of estimators in factor analysis under general conditions*, Ann. Statist. 16 (1988), pp. 759–771.

[2] J.C. Anderson and D.W. Gerbing, *The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis*, Psychometrika 49 (1984), pp. 155–173.

[3] A. Beauducel, *On the generalizability of factors: the influence of changing contexts of variables on different methods of factor extraction*, Methods Psychol. Res. Online 6 (2001), pp. 69–96.

[4] A. Boomsma, *Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation*, Psychometrika 50 (1985), pp. 229–242.

[5] N.E. Briggs and R.C. MacCallum, *Recovery of weak common factors by maximum likelihood and ordinary least squares estimation*, Multivariate Behav. Res. 38 (2003), pp. 25–56.

[6] T.A. Brown, *Confirmatory Factor Analysis for Applied Research*, The Guilford Press, New York, 2006.

[7] M.W. Browne, *A comparison of factor analytic techniques*, Psychometrika 33 (1968), pp. 267–334.

[8] R.B. Cattell and J. Jaspers, *A general plasmode for factor analytic exercises and research*, Multivariate Behav. Res. Monographs 3 (1967), pp. 1–12.

[9] R.B. Cattell and W. Sullivan, *The scientific nature of factors: A demonstration by cups of coffee*, Behav. Sci. 7 (1962), pp. 184–193.

[10] Y.-P. Chen, H.-C. Huang, and I.-P. Tu, *A new approach for selecting the number of factors*, Comput. Statist. Data Anal. 54 (2010), pp. 2990–2998.

[11] A.L. Comrey, *Common methodological problems in factor analytic studies*, J. Consulting Clin. Psychol. 46 (1978), pp. 648–659.

[12] J.M. Conway and A.I. Huffcutt, *A review and evaluation of exploratory factor analysis practices in organizational research*, Organizational Res. Methods 6 (2003), pp. 147–168.

[13] A.B. Costello and J.W. Osborne, *Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis*, Practical Assessment, Res. Eval. 10 (2005). Available at http://pareonline.net/pdf/v10n7.pdf.

[14] J.C.F. De Winter and D. Dodou, *The driver behaviour questionnaire as a predictor of accidents: A meta-analysis*, J. Safety Res. 4 (2010), pp. 463–470.

[15] J.C.F. De Winter, D. Dodou, and P.A. Wieringa, *Exploratory factor analysis with small sample sizes*, Multivariate Behav. Res. 44 (2009), pp. 147–181.

[16] C. DiStefano, M. Zhu, and D. Mîndrilã, *Understanding and using factor scores: Considerations for the applied researcher*, Practical Assessment, Res. Eval. 14 (2009). Available at http://pareonline.net/pdf/v14n20.pdf.

[17] L.R. Fabrigar, D.T. Wegener, R.C. MacCallum, and E.J. Strahan, *Evaluating the use of exploratory factor analysis in psychological research*, Psychol. Methods 4 (1999), pp. 272–299.

[18] J.L. Fava and W.F. Velicer, *The effects of underextraction in factor and component analyses*, Educational Psychol. Measurement 56 (1996), pp. 907–929.

[19] F.J. Floyd and K.F. Widaman, *Factor analysis in the development and refinement of clinical assessment instruments*, Psychol. Assessment 7 (1995), pp. 286–299.

[20] D.W. Gerbing and J.C. Anderson, *Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate specifications*, Psychometrika 52 (1987), pp. 99–111.

[21] J.F. Geweke and K.J. Singleton, *Interpreting the likelihood ratio statistic in factor models when sample size is small*, J. Amer. Statist. Assoc. 75 (1980), pp. 133–137.

[22] R.L. Gorsuch, *Factor Analysis*, 2nd ed., Lawrence Erlbaum, Hillsdale, NJ, 1983.

[23] R.L. Gorsuch, *New procedures for extension analysis in exploratory factor analysis*, Educational Psychol. Measurement 57 (1997), pp. 725–740.

[24] H.H. Harman, *Modern Factor Analysis*, University of Chicago Press, Chicago, 1976.

[25] L.K. Hansen, J. Larsen, and T. Kolenda, *Blind detection of independent dynamic components*, IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, Salt Lake City, UT, 2001, pp. 3197–3200.

[26] R.K. Henson and J.K. Roberts, *Use of exploratory factor analysis in published research: Common errors and some comment on improved practice*, Educational Psychol. Measurement 66 (2006), pp. 393–416.

[27] S. Hong, *Generating correlation matrices with model error for simulation studies in factor analysis: A combination of the Tucker–Koopman–Linn model and Wijsman's algorithm*, Behav. Res. Methods, Instruments Comput. 31 (1999), pp. 727–730.

[28] K.G. Jöreskog, *Some contributions to maximum likelihood factor analysis*, Psychometrika 32 (1967), pp. 443–482.

[29] K.G. Jöreskog, *Factor analysis and its extensions*, in *Factor analysis at 100: Historical Developments and Future Directions*, R. Cudeck and R.C. MacCallum, eds., Lawrence Erlbaum, Mahwah, NJ, 2007, pp. 47–77.

[30] A.S. Kaplunovsky, *Factor analysis in environmental studies*, HAIT J. Sci. Eng. B 2 (2005), pp. 54–94.

[31] D.N. Lawley and A.E. Maxwell, *Factor Analysis as a Statistical Method*, American Elsevier, New York, 1971.

[32] F.R. Lawrence and G.R. Hancock, *Conditions affecting integrity of a factor solution under varying degrees of overextraction*, Educational Psychol. Measurement 59 (1999), pp. 549–579.

[33] R.C. MacCallum, M.W. Browne, and L. Cai, *Factor analysis models as approximations*, in *Factor Analysis at 100: Historical Developments and Future Directions*, R. Cudeck and R.C. MacCallum eds., Lawrence Erlbaum, Mahwah, NJ, 2007, pp. 153–175.
[34] L.F. Marcus, *Program supplement to applied factor analysis in the natural sciences*, in *Applied Factor Analysis in the Natural Sciences*, R. Reyment and K.G. Jöreskog, eds., Cambridge University Press, Cambridge, 1993, pp. 289–351.
[35] C. Maté and R. Calderón, *Exploring the characteristics of rotating electric machines with factor analysis*, J. Appl. Statist. 27 (2000), pp. 991–1006.
[36] D.F. Morrison, *Multivariate Statistical Methods*, 2nd ed., McGraw-Hill, New York, NY, 1976.
[37] S.A. Mulaik, *Linear Causal Modeling with Structural Equations*, Chapman & Hall/CRC, Boca Raton, FL, 2009.
[38] B.P. O'Connor, *SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test*, Behav. Res. Methods, Instruments, Comput. 32 (2000), pp. 396–402.
[39] B.P. O'Connor, *EXTENSION: SAS, SPSS, and MATLAB programs for extension analysis*, Appl. Psychol. Measurement 25 (2001), p. 88.
[40] H. Ogasawara, *Correlations among maximum likelihood and weighted/unweighted least squares estimators in factor analysis*, Behaviormetrika 30 (2003), pp. 63–86.
[41] U.H. Olsson, T. Foss, S.V. Troye, and R.D. Howell, *The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality*, Structural Equation Modeling 7 (2000), pp. 557–595.
[42] Y. Parmet, E. Schechtman, and M. Sherman, *Factor analysis revisited – How many factors are there?*, Commun. Statist. – Simulation Comput. 39 (2010), pp. 1893–1908.
[43] S.C. Snook and R.L. Gorsuch, *Component analysis versus common factor analysis*: *A Monte Carlo study*, Psychol. Bull. 106 (1989), pp. 148–154.
[44] B. Thompson, *Exploratory and Confirmatory Factor Analysis*, American Psychological Association, Washington, DC, 2004.
[45] L.L. Thurstone, *Multiple-Factor Analysis. A Development and Expansion of the Vectors of Mind*, The University of Chicago, Chicago, IL, 1947.
[46] H.E.A. Tinsley and D.J. Tinsley, *Uses of factor analysis in counseling psychology research*, J. Counseling Psychol. 34 (1987), pp. 414–424.
[47] Transport Research Laboratory, *Safety, Security and Investigations Division* [*TRL*], *Cohort II*: *A Study of Learner and Novice Drivers*, *2001–2005* [*Computer file*]. Colchester, Essex, UK, Data Archive [distributor], July 2008, SN: 5985.
[48] L.R. Tucker, *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984), Department of the Army, Washington, DC, 1951.
[49] L.R. Tucker, R.F. Koopman, and R.L. Linn, *Evaluation of factor analytic research procedures by means of simulated correlation matrices*, Psychometrika 34 (1969), pp. 421–459.
[50] S. Unkel, N.T. Trendafilov, A. Hannachi, and I.T. Jolliffe, *Independent exploratory factor analysis with application to atmospheric science data*, J. Appl. Statist. 37 (2010), pp. 1847–1862.
[51] W.F. Velicer and D.N. Jackson, *Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure*, Multivar. Behav. Res. 25 (1990), pp. 1–28.
[52] W.F. Velicer and J.L. Fava, *Effects of variable and subject sampling on factor pattern recovery*, Psychol. Methods 3 (1998), pp. 231–251.
[53] P. Wells, S. Tong, B. Sexton, G. Grayson, and E. Jones, *Cohort II A study of learner and new drivers*, Volume 1 – Main report (Report No. 81). Department for Transport, London, 2008.
[54] K.F. Widaman, *Common factors versus components: Principals and principles, errors and misconceptions*, in *Factor Analysis at 100: Historical Developments and Future Directions*, R. Cudeck and R.C. MacCallum, eds., Lawrence Erlbaum, Mahwah, NJ, 2007, pp. 177–203.
[55] C. Ximénez, *Recovery of weak factor loadings in confirmatory factor analysis under conditions of model misspecification*, Behav. Res. Methods 41 (2009), pp. 1038–1052.