Ann. Inst. Statist. Math.
39 (1987), Part B, 443-455

PRAGMATIC TREATMENT OF IMPROPER SOLUTIONS IN FACTOR ANALYSIS*

MANABU SATO

(Received Feb. 4, 1985; revised Nov. 25, 1986)

Summary

In the application of factor analysis to empirical data, a statistical test almost always indicates more factors than researchers expect. However, if one more factor is tried to be extracted, proper solutions cannot be obtained frequently and several problems arise.

This paper investigates causes of the problems and proposes a pragmatic treatment of improper solutions. Further, some recommendations are made on the practical application of factor analysis.

1. Introduction

Consider the factor analysis model

$$x = \Lambda f + u , \tag{1.1}$$

where x is a random vector of p components representing observations, f is a random vector of k components representing common factors, u is a random vector of p components representing a unique part, which is made up of specific factors and errors, and Λ is a $p\times k$ ($p>k$) matrix of rank k representing factor loadings. It is assumed that $\mathrm{E}\{f\}=0$, $\mathrm{E}\{u\}=0$, $\mathrm{E}\{fu'\}=O$, $\mathrm{E}\{ff'\}=I$ (a unit matrix), and $\mathrm{E}\{uu'\}$, denoted by Ψ, is diagonal and positive definite.

If ΛA contains more than one nonzero elements in every column, where A is any nonsingular matrix, Λ is called *a common factor matrix* ([28]). If there exists a column in Λ whose elements become nearly zero except only one element by a suitable nonsingular rotation, the factor corresponding to this column is said to be *quasi-specific* ([30]). If an estimate $\hat{\Psi}$ of Ψ is not positive definite, it is called *a Heywood solution*

* This research was supported in part by Grand-in-Aid for Scientific Research of the Ministry of Education, Science and Culture under Contract Number 61730015 and 61530017.
Key words and phrases: Factor analysis, improper solution, iterative procedure, extended uniqueness, incomplete Procrustean rotation.

or an improper solution. The region where one or more of the unique variances are not positive, it is called *a Heywood region*.

Empirical data involve many common factors. Cattell ([5], p. 204) sorted factors into true factors, that are composed of i) factors large enough to worth dealing with, which are called main factors, and ii) trivial factors, and error factors. Researchers treat not trivial factors but main common factors. Rao [24] named the latter dominant common factors. In the application to psychological data, the number m_0 of main factors that researchers expect is usually assumed to be known.

Geweke and Singleton [9] studied the cases where loading matrices did not involve trivial factors by Monte Carlo experiments. They reported as follows: "When the usual regularity conditions are satisfied and sample size is at least 30, the asymptotic theory seems to be appropriate." However, for empirical data, a statistical test almost always indicates more factors. When the number m of factors being extracted is more than m_0, problems arise frequently.

In Section 2, some problems frequently encountered are described. In Section 3, causes of the problems are investigated. In Section 4, a treatment of improper solutions is proposed. Finally in Section 5, some recommendations are made.

2. Problems encountered frequently

2.1 *Problems*

In the application of factor analysis to empirical data, the following situation was commonly met. A solution with m_0 factors was proper and the loadings $\hat{\Lambda}_{m_0}$ were physically interpretable. A chi-square test for m_0-factor was, however, statistically significant ([13], [27]), and one tried to extract m_0+1 factors. Then the following problems (P1), (P2) and (P3) arose.

(P1) The intermediate value $\tilde{\Psi}$ of the iterative solution for Ψ frequently tended to go into the Heywood region ([15], [20]).

(P2) The element $\tilde{\psi}_*$ of $\tilde{\Psi}$ that tended to be negative depended often on initial estimates ([32]).

(P3) When the variable corresponding to $\tilde{\psi}_*$ was deleted, another element of $\tilde{\Psi}$ tended to be negative frequently ([8]).

These were pointed out since Jöreskog [15] has proposed a rapidly convergent method to obtain the maximum likelihood estimates.

To resolve (P1), Jöreskog [15] proposed to eliminate the variable corresponding to $\tilde{\psi}_*$ and carry out the iteration on the conditional correlation matrix. Further, ideas for the minres method ([12]) or Bayesian approach ([17], [19]) were proposed. None of them, however, settled (P2) and (P3).

In Jöreskog's method, the iteration could not be continued when (P1) arose. Tumura and Sato [30] have shown that Heywood solutions can be obtained by Jennrich and Robinson's method ([14]), which is a modification of Jöreskog's method in the sense of using eigenvalues of $R^{-1/2}\tilde{\Psi}R^{-1/2}$ instead of those of $\tilde{\Psi}^{-1/2}R\tilde{\Psi}^{-1/2}$, where R denotes a sample correlation matrix. When methods capable of obtaining Heywood solutions, for example, Jennrich and Robinson's method, were applied, another problem (P4) arose.

(P4) The iteration did not terminate frequently even in the Heywood region after a sufficiently large number of steps, consequently most likely diverged ([30]).

2.2 *Example*

A set of data ($m_0=3$), which consists of 810 observations of 10 variables, was originally analyzed by Maxwell ([21], p. 55). After that, analyses were carried out by Mattsson, Olsson and Rosén [20] (Jöreskog [15]), and Tumura, Fukutomi and Asoo [32]. Those results are shown in Lawley and Maxwell [18], pp. 44–46 and p. 34.

For $m=3$, the obtained loadings $\hat{\Lambda}_3$ was interpretable (Table 2.1). However, the test showed three-factor hypothesis was significant ($\chi^2=78.5>\chi^2_{.001}=42.3$).

For $m=4$, all the problems (P1), (P2), (P3) and (P4) arose. Table 2.2 shows (P1) and (P2); the 8th, the 6th or the 9th element of $\tilde{\Psi}$ tended to be negative depending on initial estimates. And solutions,

Table 2.1. Maxwell's data, $m=3$, the physically interpretable solution, varimax rotated loadings.

	$\hat{\Psi}$	I	II	III
1	.369	.653	−.204	.352
2	.616	.129	−.133	.591
3	.306	.469	−.204	.657
4	.644	.348	−.119	.469
5	.414	.726	−.082	.227
6	.772	−.122	.433	−.160
7	.336	−.279	.754	−.130
8	.641	−.039	.591	−.092
9	.675	−.118	.532	−.167
10	.622	−.028	.614	−.028

Table 2.2. Maxwell's data, $m=4$, Jöreskog's method, varimax rotated loadings.

(1)

	$\tilde{\Psi}$	I	II	III	IV
1	.382	.607	−.304	.395	−.029
2	.618	.106	−.121	.595	−.040
3	.303	.435	−.198	.682	−.060
4	.639	.342	−.080	.475	−.110
5	.364	.757	−.072	.233	−.061
6	.779	−.115	.355	−.171	.230
7	.273	−.245	.789	−.151	.148
8	.001	−.043	.370	−.056	.926
9	.691	−.114	.436	−.176	.273
10	.611	−.003	.615	−.049	.092

(2)

	$\tilde{\Psi}$	I	II	III	IV
1	.386	.609	−.272	.398	−.103
2	.619	.100	−.118	.596	−.048
3	.305	.429	−.187	.685	−.081
4	.645	.329	−.105	.482	−.054
5	.366	.753	−.074	.245	−.034
6	.001	−.091	.298	−.116	.942
7	.313	−.259	.765	−.159	.098
8	.658	−.048	.543	−.085	.194
9	.676	−.106	.526	−.185	.041
10	.616	−.019	.613	−.042	.075

(3)

	$\tilde{\Psi}$	I	II	III	IV
1	.368	.656	−.267	.345	−.107
2	.612	.133	−.124	.594	−.044
3	.305	.476	−.184	.653	−.088
4	.645	.360	−.094	.459	−.076
5	.415	.728	−.079	.218	−.081
6	.772	−.124	.427	−.168	.048
7	.311	−.274	.759	−.136	.138
8	.665	−.052	.511	−.084	.254
9	.001	−.106	.352	−.117	.922
10	.596	−.012	.629	−.046	.079

The used initial estimates are as follows:
(1) The value recommended by Jöreskog [15].
(2) The value used by Tumura et al. [32].
(3) The value recommended by Jöreskog, except $\tilde{\psi}_9 \fallingdotseq .571$ is replaced by $.25$.

Table 2.3. Maxwell's data, $\tilde{\Psi}$ obtained by deleting the variable that tends to go into the Heywood region, $m=4$, Jöreskog's method.

	(1)	(2)	(3)
1	.429	.389	.368
2	.621	.620	.610
3	.300	.303	.305
4	.642	.639	.636
5	.001	.353	.403
6	.798	( )	.773
7	.321	.257	.325
8	( )	.001	.001
9	.714	.689	( )
10	.320	.617	.561

(1) The 8th variable is deleted.
(2) The 6th variable is deleted.
(3) The 9th variable is deleted.
The used initial estimates are the values recommended by Jöreskog [15].

especially Factor IV, depended on initial estimates. Table 2.3 shows (P3); for instance, in case that the 8th variable was deleted, the 5th element of $\tilde{\Psi}$ tended to be negative, as shown (1) in this table. Table 2.4 shows (P4); in case that the iteration was continued using Jennrich and Robinson's method, it did not terminate.

Table 2.4. Maxwell's data, behavior of the iterative process $\tilde{\Psi}$, $m=4$.

(1) The value recommended by Jöreskog [15] is used for an initial estimate.

variable No. / count	1	2	3	4	5	6	7	8	9	10
initial est.	.386	.597	.385	.559	.485	.633	.411	*.547*	.571	.566
10	.361	.584	.322	.629	.430	.766	.329	*.301*	.669	.551
11	.370	.603	.313	.636	.397	.773	.300	*.144*	.686	.593
12	.390	.631	.298	.643	.346	.783	.255	*−.088*	.702	.637
20	.423	.665	.280	.649	.144	.792	.277	*−1.571*	.717	.617
30	.412	.645	.291	.645	.263	.801	.264	*−10.200*	.721	.621
40	.396	.629	.298	.644	.327	.800	.276	*−17.663*	.723	.608
50	.397	.631	···	.642	.324	.801	.278	*−26.797*	···	···
60	.398	.633	···	.641	.323	···	.279	*−36.797*	···	.607
70	···	···	···	···	.322	···	···	*−46.797*	···	···
80	···	···	···	···	···	···	.280	*−56.797*	···	···
90	···	···	.297	.640	···	···	···	*−66.797*	···	···
100	···	···	···	···	···	···	···	*−76.797*	···	···

Table 2.4. (Continued)

(2) The value applied by Tumura et al. [32] is used for an initial estimate.

count \ variable No.	1	2	3	4	5	6	7	8	9	10
initial est.	.440	.670	.280	.660	.120	*.110*	.320	.660	.680	.630
10	.434	.665	.273	.652	.169	*.084*	.316	.659	.677	.621
12	.412	.652	.281	.650	.270	*.035*	.315	.660	.676	.619
13	.378	.612	.310	.645	.382	*−.006*	.312	.656	.675	.614
20	...	.629	.300	.647	.381	*−.439*	.309	.673	.668	.616
30	.400	.654	.283	.651	.307	*−2.578*	.315	.677	.671	.618
40	.410	.659	.279	.652	.268	*−9.897*	...	.680	.672	.619
50	.409	...	.280	...	.272	*−19.897*	...	.681	...	.618
60	.407	.656	.281	...	.284	*−29.897*	...	...	...	...
70	.403	.654	.283	.651	.297	*−39.897*	.314	...	...	...
80	.400	.651	.285	...	.312	*−49.897*	...	...	...	.617
90	.396	.647	.288	.650	.326	*−59.897*	...	.680	...	...
100	.392	.643	.291	...	.340	*−69.897*	...	...	...	...

The same value as the above at least in the three places of decimal is denoted by '...'.

3. Causes of the problems

We investigate the reason why one more factor cannot be obtained frequently in spite of the result of test. Nonconvergence cases were studied ([1], [4]), however, when and why they occurred frequently in the analysis of empirical data were not discussed. Some effects of trivial factors have been ignored in the past study.

Experiment. The author wished to see if (P4) was caused by trivial factors. The structure which simulated empirical data with $m_0=2$ was treated. The first m_0+t columns ($t=1$, 2, 3 and 4) of the matrix given in Table 3.1 were used; the first two columns represent main factor loadings and the remainings trivial factors. Sample correlation matrices with sample size 100 were generated from random numbers using formula (1.1), where $\boldsymbol{f}$ and $\boldsymbol{u}$ are distributed as $N(0, I)$ and $N(0, \operatorname{diag}(I-\Lambda\Lambda'))$, respectively. An iteration terminated after the value of the likelihood function changed less than ε in absolute value and the

Table 3.1. Loading matrix used in the experiments.

$$\begin{bmatrix} .8 & .1 & .15 & .2 & -.15 & .2 \\ .75 & .1 & .2 & -.15 & .2 & -.15 \\ .7 & .2 & -.15 & .2 & -.15 & -.2 \\ .65 & .2 & -.2 & -.15 & .2 & .15 \\ .1 & .8 & .15 & .2 & .15 & .2 \\ .1 & .75 & .2 & -.15 & -.2 & -.15 \\ .2 & .7 & -.15 & .2 & .15 & -.2 \\ .2 & .65 & -.2 & -.15 & -.2 & .15 \end{bmatrix}$$

maximum norm of gradient got less than ε, where $\varepsilon=10^{-5}$.

Two ($=m_0$) factors were extracted and the result was classified as a proper solution, a Heywood solution, and a nonconvergence case. Further, the hypothesis for factor size 2 was tested. Next, 3 factors were extracted from the same correlation matrix and the result was classified. Two hundred and fifty replications were done for each t.

Table 3.2 summarizes the results; the increase of t made the tail of the distribution of the test statistics heavier for $m=m_0$, but decreased the number of proper solutions for $m=m_0+1$. It ascertained that (P4) is due to several trivial factors.

Population cases were dealt with in order to investigate the causes. Suppose Λ satisfied a sufficient condition for the extended uniqueness described in Appendix. If $k+s$ factors were extracted from a population covariance matrix $\Sigma=\Lambda\Lambda'+\mathrm{diag}(I-\Lambda\Lambda')$ by iterative procedures; (1) the obtained solution consisted of the true common factors and s specific factors, and (2) the variables to which specific factors were added and those loadings depended on initial estimates ([32], Experiment

Table 3.2. Results of the extraction and the test for $m=m_0$, and the extraction for $m=m_0+1$; when t trivial factors exist (Monte Carlo experiments with 250 replications for each t).

		t	1	2	3	4
$m=m_0$	proper solution		100.0%	99.6%	99.6%	99.2%
	Heywood solution		.0	.4	.4	.8
	nonconvergence		.0	.0	.0	.0
	-1% level*		19.2	39.2	55.2	52.8
	-5		42.8	66.0	78.0	80.0
	-25		76.8	90.8	96.8	95.6
	-50		88.8	97.6	99.6	99.2
	-100		100.0	100.0	100.0	100.0
$m=m_0+1$	proper solution		60.8	53.2	20.4	16.8
	Heywood solution		20.8	27.2	38.8	14.4
	nonconvergence		18.4	19.6	40.8	68.8
#	$m=m_0$	$m=m_0+1$				
	proper	proper	59.3	54.5	17.4	14.5
	proper	Heywood	25.0	28.5	42.1	15.5
	proper	nonconv.	15.7	16.4	40.0	69.0
	Heywood	proper	.0	.0	.0	.0
	Heywood	Heywood	.0	.0	.5	.0
	Heywood	nonconv.	.0	.6	.0	1.0

Results of the extraction for $m=m_0+1$ from the samples whose χ^2 values for $m=m_0$ are significant at the 5% level.

* Frequency of the test statistics corresponding to the upper probabilities.

1). Problems (P1), (P2) and (P3) arose even in the population cases, which did not have unfitness for a model and sampling fluctuation. Sample covariance matrices were not exactly decomposed as (A.1) in probability one. Therefore quasi-specific factors appeared and (P4) arose. Nonconvergence cases ([1], [4]) or Heywood solutions ([6], [7], [25]) were investigated, however, this interpretation was not pointed out.

4. Treatment of improper solutions

4.1 *Proposed treatment*

For $m=m_0+1$, apply methods capable of obtaining Heywood solutions, and continue the iteration until all elements of $\tilde{\Psi}$ except one element are stable. After that, rotate the indeterminate loadings $\tilde{\Lambda}_{m_0+1}$ to satisfy

$$\tilde{\Lambda}_{m_0+1}T \doteqdot [\hat{\Lambda}_{m_0}; s] , \tag{4.1}$$

where T is an orthogonal matrix of order m_0+1 and s represents specific factor loadings. If (4.1) holds, interpret $\hat{\Lambda}_{m_0}$.

For this rotation, the following iterative procedure is recommended:

Step 0: Arrange columns of $\tilde{\Lambda}_{m_0+1}$ so that the first m_0 columns consist of main factor loadings and the last quasi-specific ones. Find the position of the maximum absolute value in the last column; say (i, m_0+1). Set $s'=(0, 0, \cdots, 0)$.

Step 1: Set (i, m_0+1) element of the current loading matrix to the i-th element of s.

Step 2: Rotate the loading matrix by

$$T=(\Lambda^{*\prime}\Lambda^0\Lambda^{0\prime}\Lambda^*)^{-1/2}\Lambda^{*\prime}\Lambda^0 , \tag{4.2}$$

where Λ^* is the current loading matrix and $\Lambda^0=[\hat{\Lambda}_{m_0}; s]$. If every loading changes less than δ, a small positive constant for a convergence criterion, by the rotation, then stop; otherwise, return to Step 1.

4.2 *Bases of the treatment*

Unless such methods are applied and the iteration is continued, one cannot find out whether quasi-specific factors appear. Further the iteration may terminate in the proper region passing through the Heywood region ([30], Example 1).

The rotation matrix T is derived as follows: Consider the orthogonal matrix T to fit a matrix Λ^* to a target matrix Λ^0. Suppose that the matrices Λ^* and Λ^0 have the same order and are of full rank. The least squares fit is given by formula (4.2) ([10]), if a matrix $\Lambda^{*\prime}\Lambda^0$ is nonsingular. Since experimenters possess information only about the position of the nonzero element of s, the use of an iterative procedure, which is called the incomplete Procrustean rotation ([26]), is proposed. When δ is set to 10^{-6}, at most ten iterations are required to converge from author's experience.

The treatment worked well for many sets of empirical data; for example, the data dealt by Mattsson, Olsson and Rosén [20] (Bechtoldt ([3], p. 412, Sample 1), Harman ([11], p. 82 and 137), Maxwell ([21], p. 55) and Rao ([23], p. 110)), Jöreskog ([16], p. 152, Problem 2) and Maxwell ([22], p. 9).

On the other hand, Driel, Prins and Veltkamp [7] obtained a complex solution, a solution such as $\hat{\Sigma}-\hat{\Psi}$ has at most $m-1$ positive eigenvalues. In such cases, they recommended to delete the variables which had large asymptotic variances ([6], [7]), however, it did not settle the problems. By a suitable rotation, it is found out very often that the loadings of complex solutions consist of two parts; (1) the real number loadings that quite agree with $\hat{\Lambda}_{m_0}$, and (2) the real or imaginary number loadings that represent trivial factors or quasi-specific factors.

4.3 *Example*

The data dealt with in subsection 2.2 were analyzed.

If the iteration was continued using Jennrich and Robinson's method; the loadings corresponding to Factors I, II and III were nearly invariant, and, the only one loading increased and the remainings decreased in absolute value in the column corresponding to Factor IV, therefore, it became a quasi-specific factor. Table 2.4 shows the behavior of the iterative process. Remark that the main factor loadings closely agreed with $\hat{\Lambda}_3$ independently initial estimates ((1), (2) and (3) in Table 4.1, cf.

Table 4.1. Maxwell's data, at the 100th iterative count, the rotated loadings fitting to $\hat{\Lambda}_3$ and specific factor loadings.

(1)

	$\tilde{\Psi}$	I	II	III	IV
1	.398	.608	−.302	.376	.006
2	.632	.134	−.140	.575	.002
3	.298	.452	−.213	.673	.002
4	.643	.355	−.106	.469	−.005
5	.322	.797	−.060	.198	−.005
6	.802	−.130	.393	−.164	.012
7	.279	−.271	.796	−.122	−.005
8	−76.805	−.039	.591	−.092	8.800
9	.724	−.133	.480	−.168	.015
10	.607	−.030	.626	−.019	−.006

(2)

	$\tilde{\Psi}$	I	II	III	IV
1	.395	.620	−.287	.372	−.003
2	.637	.138	−.134	.570	.001
3	.291	.455	−.205	.679	.000
4	.648	.352	−.118	.463	.001
5	.342	.782	−.074	.200	.002
6	−69.899	−.122	.433	−.160	8.407
7	.314	−.267	.772	−.136	−.003
8	.678	−.066	.558	−.076	.013
9	.675	−.116	.532	−.169	−.008
10	.617	−.028	.617	−.026	−.002

Table 4.1. (Continued)

(3)

	$\tilde{\Psi}$	I	II	III	IV
1	.357	.667	−.277	.348	−.004
2	.589	.114	−.137	.616	.003
3	.321	.476	−.205	.641	−.000
4	.648	.354	−.115	.462	−.001
5	.444	.701	−.086	.241	.002
6	.776	−.124	.427	−.162	−.009
7	.293	−.280	.783	−.120	−.005
8	.694	−.055	.542	−.090	.017
9	−57.985	−.118	.532	−.167	7.659
10	.607	−.014	.626	−.039	−.004

(4)

	$\tilde{\Psi}$	I	II	III	IV
1	.398	.612	−.294	.376	.006
2	.627	.131	−.138	.580	.002
3	.300	.454	−.211	.671	.002
4	.644	.357	−.106	.466	−.006
5	.329	.792	−.059	.201	−.006
6			*(deleted)*		
7	.254	−.273	.811	−.119	−.007
8	−62.789	−.039	.591	−.092	7.964
9	.727	−.134	.474	−.171	.016
10	.621	−.031	.614	−.025	−.006

The used initial estimates are as follows:
(1) and (4) ··· The value recommended by Jöreskog [15].
(2) ··· The value used by Tumura et al. [32].
(3) ··· The value recommended by Jöreskog, except $\tilde{\psi}_9 \fallingdotseq .571$ is replaced by $.25$.

Table 4.2. Maxwell's data, $m=4$, the complex solution, the rotated loadings fitting the real number loadings to $\hat{\Lambda}_3$.

	$\hat{\Psi}$	$\hat{\sigma}_\psi$	I	II	III	IV
1	.399	.036	.607	−.304	.375	*.038i*
2	.628	.058	.131	−.139	.579	*−.001i*
3	.299	.040	.453	−.216	.670	*.009i*
4	.645	.038	.355	−.133	.466	*−.029i*
5	.329	.103	.792	−.071	.199	*−.011i*
6	.813	.047	−.125	.400	−.164	*.125i*
7	.275	.044	−.262	.802	−.122	*.036i*
8	2.179	2.296	−.089	.465	−.086	*−1.188i*
9	.740	.048	−.127	.487	−.168	*.148i*
10	.610	.045	−.021	.624	−.023	*.018i*

($i^2=-1$)

Table 2.1). Further, even if the variable corresponding to $\tilde{\psi}_*$ was deleted, the above tendency persisted; for instance, the result of deletion of the 6th variable is shown in (4) in Table 4.1. Thus, actually, the main common factors were determined, and consequently, all the problems (P1), (P2), (P3) and (P4) were overcome simultaneously.

As for the complex solution obtained by Driel, Prins and Veltkamp ([7], (5) in Table 4), the real number loadings quite agreed with $\hat{\Lambda}_3$ by a rotation (Table 4.2, cf. Table 2.1).

5. Recommendations

If a uniqueness condition of Λ is violated, samples from such population yield improper solutions very often ([6]). A necessary condition for uniqueness is that each column of ΛA has at least three nonzero elements for every nonsingular matrix A ([2], Theorem 5.6).

The statistical test almost always indicates more factors in the analysis of empirical data. Hence, various goodness-of-fit indices were proposed and examined ([1], [27]). However, they were not taken account of the uniqueness condition and an existence of trivial factors. When we decide factor size, we should count factors that have a great influence on at least three variables.

After deleting one or more of the variables from the original correlation matrix taking into consideration of the uniqueness condition, check whether similar solutions can be obtained. Unless solutions are relatively consistent under the selection of test batteries, results are unreliable ([8]).

When we make test batteries, it is to be desired that a hypothetical loading matrix should satisfy *the extended uniqueness condition.* It is required in view of the following situations:

1) Some variables are deleted in order to check the stability of solutions.

2) Extraction of one more factor is often tried as a result of the test.

Acknowledgements

The author would like to express his sincere gratitude to Professor Yosiro Tumura for his instructive suggestions and discussions throughout this study, and wishes to thank Dr. Kazuo Fukutomi for encouragement. He is also indebted to referees for valuable comments.

SCIENCE UNIVERSITY OF TOKYO

REFERENCES

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

[2] Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis, *Proc. Third Berkeley Symp. Math. Statist. Prob.*, **5**, 111-150.

[3] Bechtoldt, H. P. (1961). An empirical study of the factor analysis stability hypothesis, *Psychometrika*, **26**, 405-432.

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

[5] Cattell, R. B. (1965). Factor analysis: an introduction to essentials, *Biometrics*, **21**, 190-215.

[6] Driel, O. P. van (1978). On various causes of improper solutions in maximum likelihood factor analysis, *Psychometrika*, **43**, 225-243.

[7] Driel, O. P. van, Prins, H. J. and Veltkamp, G. W. (1974). Estimating the parameters of the factor analysis model without the usual constraints of positive definitness, *Proceedings of the Symposium on Computational Statistics* (*COMPSTAT*), Vienna, 255-265.

[8] Fukutomi, K. (1973). On the adequacy of factor extraction, *TRU Mathematics*, **9**, 119-136.

[9] Geweke, J. F. and Singleton, K. J. (1980). Interpreting the likelihood ratio statistics in factor models when sample size is small, *J. Amer. Statist. Ass.*, **75**, 133-137.

[10] Green, B. F. (1952). The orthogonal approximation of an oblique structure in factor analysis, *Psychometrika*, **17**, 429-440.

[11] Harman, H. H. (1960). *Modern Factor Analysis*, University of Chicago Press, Chicago.

[12] Harman, H. H. and Fukuda, Y. (1966). Resolution of the Heywood case in the minres solution, *Psychometrika*, **31**, 563-571.

[13] Jackson, D. N. and Chan, D. W. (1980). Maximum likelihood estimation in common factor analysis: A causionary note, *Psychol. Bull.*, 88, 502-508.

[14] Jennrich, R. I. and Robinson, S. M. (1969). A Newton-Raphson algorithm for maximum likelihood factor analysis, *Psychometrika*, **34**, 111-123.

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

[16] Jöreskog, K. G. (1977). Factor analysis by least-squares and maximum-likelihood methods (eds. Enslein, K., Ralston, A. and Wilf, H. S.), *Statistical Methods for Digital Computers*, Wiley-Interscience, New York, 125-153.

[17] Koopman, R. F. (1978). On Bayesian estimation in unrestricted factor analysis, *Psychometrika*, **43**, 109-110.

[18] Lawley, D. N. and Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method*, 2nd ed., Butterworths, London.

[19] Martin, J. K. and McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: a treatment for Heywood cases, *Psychometrika*, **40**, 505-517.

[20] Mattsson, A., Olsson, U. and Rosén, M. (1966). The maximum likelihood method in factor analysis with special consideration to the problem of improper solutions, *Research Report*, Institute of Statistics, University of Uppsala, Sweden.

[21] Maxwell, A. E. (1961). Recent trends in factor analysis, *J. R. Statist. Soc.*, A124, 49-59.

[22] Maxwell, A. E. (1972). Factor analysis: Thomson's sampling theory recalled, *Brit. J. Math. Statist. Psychol.*, **25**, 1-21.

[23] Rao, C. R. (1955). Estimation and tests of significance in factor analysis, *Psychometrika*, **20**, 93-111.

[24] Rao, C. R. (1965). *Linear Statistical Inference and Its Application*, Wiley, New York.

[25] Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems, *Sociol. Meth. Res.*, **13**, 109-119.

[26] Shiba, S. (1972). *Factor Analysis* (in Japanese), University of Tokyo Press, Tokyo.

[27] Tucker, L. R. and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis, *Psychometrika*, **38**, 1-10.

[28] Tumura, Y. and Fukutomi, K. (1968). On the identification in factor analysis, *Rep. Statist. Appl. Res.*, **15**, 98-103.

[29] Tumura, Y. and Sato, M. (1980). On the identification in factor analysis, *TRU Mathematics*, **16**(2), 121-131.

[30] Tumura, Y. and Sato, M. (1981). On the convergence of iterative procedures in factor analysis, *TRU Mathematics*, **17**, 159-168.

[31] Tumura, Y. and Sato, M. (1981). On the convergence of iterative procedures in factor analysis (II), *TRU Mathematics*, **17**, 255-271.

[32] Tumura, Y., Fukutomi, K. and Asoo, Y. (1968). On the unique convergence of iterative procedures in factor analysis, *TRU Mathematics*, **4**, 52-59.

Appendix

THEOREM. (A sufficient condition for the extended uniqueness ([28], [29])) *Let*

$$\Sigma = \Lambda_k \Lambda_k' + \Psi_k$$

hold, where Λ_k *is a* $p \times k$ *matrix* ($p > k$) *and* Ψ_k *is a diagonal matrix. Suppose that there remain two disjoint submatrices of rank* k *in* Λ_k *after deletion of any* $r+1$ *rows of* Λ_k.

If Σ *has another decomposition such that*

$$\Sigma = \Lambda_{k+s} \Lambda_{k+s}' + \Psi_{k+s} \tag{A.1}$$

where Λ_{k+s}: $p \times (k+s)$, *rank* $\Lambda_{k+s} = k+s$, $0 \leqq s \leqq r$, *r and s are non-negative integers,*
Ψ_{k+s}: *a diagonal matrix,*

then, there exists an orthogonal matrix T_{k+s} *such that* $\Lambda_{k+s} T_{k+s} = [\Lambda_k ; S_s]$ *where* off-diag $S_s S_s' = \boldsymbol{O}$, *namely,* $S_s (p \times s)$ *represents specific factor loadings.*

Under the condition of this theorem, even if factor size is increased up to $k+s$, no common factor but s specific factors are added, and, the common factor matrix remains invariant. This theorem reduces to Anderson and Rubin's ([2], Theorem 5.1), when $r=0$.