Computational Statistics and Data Analysis 99 (2016) 189–203

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A new method for simultaneous estimation of the factor model parameters, factor scores, and unique parts

Alwin Stegeman *

Heymans Institute for Psychological Research, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 14 July 2015
Received in revised form 10 January 2016
Accepted 20 January 2016
Available online 10 February 2016

Keywords:
Common factor analysis
Simultaneous estimation
Direct fitting
Minimum rank factor analysis
Factor indeterminacy

ABSTRACT

In the common factor model the observed data is conceptually split into a common covariance producing part and an uncorrelated unique part. The common factor model is fitted to the data itself and a new method is introduced for the simultaneous estimation of loadings, unique variances, factor scores, and unique parts. The method is based on Minimum Rank Factor Analysis and allows for the percentage of explained common variance to be computed. Taking into account factor indeterminacy, an explicit description of the complete class of solutions for the factor scores and unique parts is given. The method is evaluated in a simulation study and fitted to a dataset in the literature.

1. Introduction

We consider exploratory factor analysis of continuous observed variables. In common factor analysis the observed variables are conceptually split up into a common part and a unique part. The common part of each variable is correlated to other observed variables, while the unique part is not. The latter includes measurement error and possibly a specific part uniquely measured by the corresponding observed variable. The common part of all observed variables is approximated by a small number of underlying latent factors. The origins of the common factor model (or factor analysis model) date back to Spearman (1904) and Thurstone (1935). A distinction is made in the literature between the *random factor model*, in which the factors and unique parts are considered random variables, and the *fixed factor model*, in which only the unique parts are random variables and the factor scores are parameters to be estimated. Fitting of the factor model is commonly done on the observed covariance or correlation matrix, with the loadings, unique variances, and factor correlations as parameters to be estimated. For the fixed factor model the factor scores are commonly estimated in a second step, using a (weighted) least squares criterion.

In this paper, we introduce a factor model that addresses two problematic issues in the foundation and application of common factor analysis. The first issue concerns the measures that are used to assess the fit of the factor model. In practice, the fit is assessed by comparing the observed correlation matrix to that of the estimated factor model, including the unique variances. This is done by, e.g., least squares (Harman and Jones, 1966), or a chi-square measure in a maximum likelihood framework (Jöreskog, 1967). However, no distinction is made between the common variances to be explained and the common variances in the estimated factor model. The latter are produced as "communalities" by statistical software packages, thus suggesting that 100% of the common variance is explained. But in practice we have imperfect fit of the factor

* Tel.: +31 50 363 6193; fax: +31 50 363 6304.
E-mail address: a.w.stegeman@rug.nl.
URL: http://www.alwinstegeman.nl.

http://dx.doi.org/10.1016/j.csda.2016.01.012
0167-9473/

model and not all common variance will be explained. The question remains how well the latent factors explain the common variance in the observed variables, which is important when choosing the number of factors. The only estimation method that provides an answer to this question and yields a percentage of explained common variance is Minimum Rank Factor Analysis (MRFA) of Ten Berge and Kiers (1991). See also Ten Berge (1998).

The second issue concerns the nonuniqueness of factor scores under the factor model, known in the literature as *factor indeterminacy* and first described by Wilson (1928). Factor indeterminacy is a fundamental property of the common factor model and occurs because the model contains more factors, including unique parts, than observed variables. Guttman (1955) showed that under perfect fit the factor scores (and unique parts) can be written as the sum of a determinate part and an orthogonal indeterminate part. The determinate part is the regression of the vector of factor scores (or the unique part) on the observed variables, while the indeterminate part represents all possible residuals of this regression. Moreover, Guttman (1955) proved a tight lower bound on the correlation between alternative factors (with the same determinate but different indeterminate parts). When this minimal correlation is close to zero, the indeterminacy of the factor scores also affects the interpretation of the factor. Indeed, how could nearly uncorrelated factors represent the same latent trait? Although factor indeterminacy may have serious consequences for factor interpretation and factor score estimation, it is usually ignored in applications and minimal correlations are not reported (Steiger, 1979; Maraun, 1996). In the empirical Bayes framework for factor score estimation (which implies using unweighted least squares) the uncertainty in the factor scores due to indeterminacy can be computed explicitly as the posterior variance of the factor scores given the observed data and estimates of the loadings, unique variances, and factor correlations (Bartholomew, 1981; Skrondal and Rabe-Hesketh, 2004). However, this implies that factors are treated as random variables which is seen by some as contradicting the idea that factor scores are parameters to be estimated (Bartholomew, 1981). In the literature, factor indeterminacy remains a highly controversial topic and experts do not seem to agree on how to address the issue (Maraun, 1996). For an overview see also Mulaik (2010).

In this paper, we present a novel model and estimation procedure of common factor analysis in which the loadings, unique variances, factor scores, and unique parts are all parameters and are estimated simultaneously. We refer to this type of factor model as the *data factor model*. Our model is a constrained version of the data factor model introduced independently by Kiers in Sočan (2003) and by De Leeuw (2004), and combines MRFA with unweighted least squares estimation of factor scores and unique parts. As a result, the explained common variance can be computed. Moreover, we extend the analysis of factor indeterminacy of Guttman (1955) to the case of imperfect fit and apply it to our model and that of De Leeuw (2004). We obtain a description of the complete class of solutions for the factor scores and unique parts, decomposed into determinate and indeterminate parts, and an expression for the minimal correlation between alternative factors. The unweighted least squares factor score estimate, its sampling variance, and its variance due to indeterminacy equal their empirical Bayes counterparts. The description of the complete class of solutions for the factor scores makes it possible to obtain probability densities of factor scores due to indeterminacy by sampling the indeterminate parts. The sampling is done uniformly and without assuming a distributional form for the factor scores. These probability densities can be used in practice as a visual tool to aid the researcher in deciding how to value the obtained factor score estimates in the presence of indeterminacy. In the empirical Bayes framework such probability densities can also be obtained, but only after specifying a prior distribution for the factor scores (Bartholomew, 1981).

Large sample theory under normality has yet been derived for the data factor model of Sočan (2003) and De Leeuw (2004), but the asymptotic distribution of the unique variance estimates by MRFA can be found in Shapiro and Ten Berge (2002). However, finite sample standard errors for the estimated loadings, unique variances, and factor correlations under the data factor models can be obtained via a bootstrap procedure (e.g., Zhang, 2014).

As a byproduct of our analysis of factor indeterminacy, we provide a mathematical explanation of the empirically observed finding of Schönemann and Wang (1972) and Grice (2001) that factors with less explained (common) variance tend to have smaller minimal correlations. Moreover, together with similar but partial observations by Bartholomew (1981) and Skrondal and Rabe-Hesketh (2004), our analysis explicitly bridges the gap that some observe between the empirical Bayes approach in which factor scores are treated as random, and the fixed factor model in which factor scores are parameters to be estimated.

The paper is organized as follows. In Section 2, we describe the factor models formally. In Section 3 we provide a brief discussion of factor indeterminacy, including some results for imperfect fit. In Section 4 we formulate the assumptions of the data factor model of De Leeuw (2004) (model I) and our constrained version (model II), derive algorithms for the simultaneous estimation of their parameters, and give expressions for the determinate and indeterminate parts of the estimates of the factor scores and the unique parts. Section 5 contains a simulation study in which we compare the performance of the algorithms for models I and II to the existing MINRES method for the random factor model. In Section 6 we fit our model II to a dataset in the literature and demonstrate its practical merits. Finally, Section 7 contains a discussion of our findings.

2. Factor model descriptions

We use the following notation. We write scalars, column vectors, and matrices as z, $\mathbf{z}$, and $\mathbf{Z}$ respectively. The size of a $p \times q$ matrix $\mathbf{Z}$ is specified as $\mathbf{Z} \in \mathbb{R}^{p\times q}$. The transpose is denoted as $\mathbf{z}^T$, the inverse is denoted as $\mathbf{Z}^{-1}$, and we use $\mathbf{Z}^{-T} = (\mathbf{Z}^T)^{-1} = (\mathbf{Z}^{-1})^T$. The $p\times p$ identity matrix is denoted by $\mathbf{I}_p$, a zero matrix is denoted by $\mathbf{O}$, a zero vector is denoted by $\mathbf{0}$,

and the $p \times 1$ vector containing only ones is denoted by $\mathbf{1}_p$. The sum of squares of a matrix is denoted as $\text{ssq}(\mathbf{Z}) = \text{trace}(\mathbf{Z}^T\mathbf{Z})$. We denote the space spanned by the columns of $\mathbf{Z}$ as $\text{span}(\mathbf{Z})$, and its orthogonal complement as $\text{span}(\mathbf{Z})^\perp$. We use $\text{diag}(\mathbf{Z})$ to denote the diagonal matrix containing the diagonal of (square) $\mathbf{Z}$ as its diagonal. Finally, an estimate of $\mathbf{Z}$ is denoted by $\widehat{\mathbf{Z}}$.

In general, the common factor model can be written as

$$X_j = \sum_{r=1}^{R} p_{jr} F_r + u_j E_j, \quad j = 1, \ldots, J, \tag{1}$$

where X_j is observed variable j, F_r is factor r, p_{jr} is the loading of variable j on factor r, E_j is the unique part of variable j, and u_j is the unique standard deviation. In the sequel it is assumed that X_j, F_r, and E_j are standardized. Although possibly deviating from the usual notation, we write the unique part as $u_j E_j$ to distinguish between the estimates of the unique variances and the standardized unique parts in the data factor models.

In the random factor model both F_r and E_j are random variables in the population of interest, and it is assumed that $\text{corr}(F_r, E_j) = 0$ for all r, j, and $\text{corr}(E_j, E_k) = 0$ for $j \neq k$. Let $\mathbf{\Phi} \in \mathbb{R}^{R\times R}$ denote the factor correlation matrix, $\mathbf{P} \in \mathbb{R}^{J\times R}$ the loading matrix, $\mathbf{U} \in \mathbb{R}^{J\times J}$ the diagonal matrix with unique standard deviations. The correlation model corresponding to (1) is

$$\mathbf{\Sigma} = \mathbf{P}\,\mathbf{\Phi}\,\mathbf{P}^T + \mathbf{U}^2, \tag{2}$$

where $\mathbf{\Sigma} \in \mathbb{R}^{J\times J}$ denotes the correlation matrix of the observed variables X_j in the population of interest. Matrix $\mathbf{P\Phi}$ contains the correlations between observed variables and factors, and is used to interpret the common factors in terms of the observed variables. When $\mathbf{\Phi} = \mathbf{I}_R$, the factors are uncorrelated or orthogonal. Correlated factors are called oblique. To obtain an interpretable solution with simple structure in $\mathbf{P\Phi}$, a rotation $\mathbf{T}$ can be applied that leaves the fitted common covariance part invariant: $\mathbf{P\Phi P}^T = (\mathbf{PT})(\mathbf{T}^{-1}\mathbf{\Phi}\mathbf{T}^{-T})(\mathbf{PT})^T$. For an overview of rotation methods, see e.g. Browne (2001). Estimating $\mathbf{P}$, $\mathbf{\Phi}$, and $\mathbf{U}$ is done by fitting (2) to the observed correlation matrix $\mathbf{S} \in \mathbb{R}^{J\times J}$. Commonly used procedures include least squares minimization of the residual (MINRES) of Harman and Jones (1966), where the residual is the difference between $\mathbf{S}$ and the model (2), and maximum likelihood factor analysis (MLFA) of Jöreskog (1967).

Minimum Rank Factor Analysis (MRFA) of Ten Berge and Kiers (1991) computes $\mathbf{U}^2$ such that the amount of unexplained variance in the approximation $\mathbf{S} - \mathbf{U}^2 \approx \mathbf{P\Phi P}^T$ is minimal and $\mathbf{S} - \mathbf{U}^2$ is a covariance matrix (i.e., positive semi-definite, has nonnegative eigenvalues). The latter is usually not the case in MINRES and MLFA and makes it meaningful to compute the percentage of explained common variance as $100 \cdot \text{trace}(\mathbf{P\Phi P}^T)/\text{trace}(\mathbf{S} - \mathbf{U}^2)$. Now $\text{trace}(\mathbf{S} - \mathbf{U}^2)$ is the total common variance to be explained and $\text{trace}(\mathbf{P\Phi P}^T)$ is the explained common variance.

The fixed factor model for subject i can be written as

$$X_{ij} = \sum_{r=1}^{R} p_{jr} f_{ir} + u_j E_j, \quad j = 1, \ldots, J, \; i = 1, \ldots, N, \tag{3}$$

where f_{ir} is the factor score of subject i on factor r. Let $\mathbf{X} \in \mathbb{R}^{N\times J}$ contain the observed data, and $\mathbf{F} \in \mathbb{R}^{N\times R}$ be the matrix of factor scores. McDonald (1979) showed that simultaneous maximum likelihood estimation of $\mathbf{P}$, $\mathbf{U}$, $\mathbf{F}$, and $\mathbf{\Phi}$ is impossible due to unboundedness. However, simultaneous estimation is possible by maximizing the likelihood ratio with respect to the saturated model. In that case, the estimates of $\mathbf{P}$, $\mathbf{U}$, and $\mathbf{\Phi}$ are identical to the corresponding maximum likelihood estimates under the random factor model (McDonald, 1979). The estimates of $\mathbf{F}$, however, are not unique and can be written as a determinate part plus an orthogonal indeterminate part analogous to Guttman (1955).

In practice, estimates of factor scores are usually obtained in a two-step procedure. First, estimates of $\mathbf{P}$, $\mathbf{U}$, and $\mathbf{\Phi}$ are obtained under the random factor model. Second, an estimate of $\mathbf{F}$ is obtained as $\widehat{\mathbf{F}} = \mathbf{X}\mathbf{B}$ for some weights matrix $\mathbf{B} \in \mathbb{R}^{J\times R}$; see Grice (2001) for an overview. The popular least squares estimate of $\mathbf{F}$ is given by $\widehat{\mathbf{F}} = \mathbf{X}\mathbf{S}^{-1}\widehat{\mathbf{P}}\,\widehat{\mathbf{\Phi}}$ and was first proposed by Thurstone (1935). It is identical to the so-called empirical Bayes estimate of $\mathbf{F}$, which is the posterior mean of $\mathbf{F}$ given the data $\mathbf{X}$ and the estimates of $\mathbf{P}$, $\mathbf{U}$, and $\mathbf{\Phi}$, which are treated as fixed (Skrondal and Rabe-Hesketh, 2004).

In this paper, we consider the data factor model in which also the unique parts are treated as fixed parameters. We write it as

$$\mathbf{X} = \mathbf{F}\,\mathbf{P}^T + \mathbf{E}\,\mathbf{U}, \tag{4}$$

where $\mathbf{E}$ denotes the $N \times J$ matrix with the unique parts. For subject i and variable j the model (4) is $x_{ij} = \sum_{r=1}^{R} p_{jr} f_{ir} + u_j e_{ij}$, which shows the relation to the fixed factor model (3). We consider simultaneous estimation of $\mathbf{F}$, $\mathbf{P}$, $\mathbf{E}$, and $\mathbf{U}$ by minimizing the sum of squares of $\mathbf{X} - \mathbf{F}\,\mathbf{P}^T - \mathbf{E}\,\mathbf{U}$.

3. Factor indeterminacy

Here, we discuss and elaborate upon some classic results on factor indeterminacy that we need in Section 4. Guttman (1955) has shown that when the factor model written as (4) fits perfectly both $\mathbf{F}$ and $\mathbf{E}$ can be written as the sum of a determinate part and an indeterminate part. The determinate parts are the regressions of $\mathbf{F}$ and $\mathbf{E}$, respectively, on the data $\mathbf{X}$ and the indeterminate parts represent all possible residuals of the corresponding regressions. Formally, Guttman (1955)

has shown that for fixed $\mathbf{P}$, $\mathbf{\Phi}$, and $\mathbf{U}$, all solutions $(\mathbf{F}, \mathbf{E})$ of (4) under the constraints $N^{-1}\mathbf{F}^T\mathbf{E} = \mathbf{O}$, $N^{-1}\mathbf{F}^T\mathbf{F} = \mathbf{\Phi}$, and $N^{-1}\mathbf{E}^T\mathbf{E} = \mathbf{I}_J$ are given by

$$\mathbf{F} = \mathbf{X}\mathbf{S}^{-1}\mathbf{P}\mathbf{\Phi} + \mathbf{V}\mathbf{W}^T, \tag{5}$$

$$\mathbf{E} = \mathbf{X}\mathbf{S}^{-1}\mathbf{U} \quad - \mathbf{V}\mathbf{W}^T\mathbf{P}^T\mathbf{U}^{-1}, \tag{6}$$

where $\mathbf{S} = N^{-1}\mathbf{X}^T\mathbf{X}$ is the data correlation matrix, and $\mathbf{V} \in \mathbb{R}^{N\times R}$ and $\mathbf{W} \in \mathbb{R}^{R\times R}$ satisfy

$$N^{-1}\mathbf{V}^T\mathbf{V} = \mathbf{I}_R, \qquad N^{-1}\mathbf{V}^T\mathbf{X} = \mathbf{O}, \qquad \mathbf{W}\mathbf{W}^T = \mathbf{\Phi} - (\mathbf{P}\mathbf{\Phi})^T\mathbf{S}^{-1}(\mathbf{P}\mathbf{\Phi}). \tag{7}$$

The determinate parts of $\mathbf{F}$ and $\mathbf{E}$ are the first terms on the right-hand sides of (5) and (6), and it can be verified that they are indeed the regression of $\mathbf{F}$ and $\mathbf{E}$ on $\mathbf{X}$, respectively. The indeterminate parts are the second terms on the right-hand sides of (5) and (6), and form the residuals of the regressions. The indeterminate parts feature matrix $\mathbf{V}$, which is arbitrary under the constraints in (7). The columns of $\mathbf{V}$ form an orthonormal basis for the residual column space, while $\mathbf{W} = N^{-1}\mathbf{F}^T\mathbf{V}$ contains the correlations between $\mathbf{F}$ and $\mathbf{V}$.

One could argue that factor indeterminacy does not affect factor interpretation since the correlations between the variables in $\mathbf{X}$ and the factors in $\mathbf{F}$ are given by $\mathbf{P}\mathbf{\Phi}$ which is usually unique. However, Guttman (1955) has shown that the correlation between two alternative rth factors $\mathbf{f}_r$ and $\widetilde{\mathbf{f}}_r$ with the same determinate but different indeterminate parts, has a tight lower bound:

$$\mathrm{Corr}(\mathbf{f}_r, \widetilde{\mathbf{f}}_r) \geq 2R_r^2 - 1, \tag{8}$$

where R_r^2 is the squared multiple correlation corresponding to the regression of $\mathbf{f}_r$ on the data $\mathbf{X}$, and can be computed as $R_r^2 = ((\mathbf{P}\mathbf{\Phi})^T\mathbf{S}^{-1}(\mathbf{P}\mathbf{\Phi}))_{rr}$. The correlation between two maximally different factors $\mathbf{f}_r$ and $\widetilde{\mathbf{f}}_r$ is equal to the lower bound $2R_r^2 - 1$, and may be small or even negative. This would make interpretation of the factor a dubious affair. As more factors are added to the model, the minimal correlation of the weakest factor tends to become smaller (Schönemann and Wang, 1972; Grice, 2001). Hence, there seems to be a trade-off between more explained (common) variance and less factor indeterminacy, which we explain mathematically at the end of this section.

The analysis of Guttman (1955) is valid for perfect fit. In the lemma below, we derive some results on factor indeterminacy under imperfect fit. We need these results in Section 4 where we derive the algorithm for the data factor model II.

Lemma 3.1. *Let $\mathbf{F} = \mathbf{F}^{(d)} + \mathbf{F}^{(u)}$ be a decomposition of mean-zero and unit-variance factors $\mathbf{F}$ into a determinate part and an indeterminate part, with $N^{-1}\mathbf{F}^T\mathbf{F} = \mathbf{\Phi}$, $\mathbf{F}^{(d)} \in \mathrm{span}(\mathbf{X})$, and $\mathbf{F}^{(u)} \in \mathrm{span}(\mathbf{X})^\perp$, where $\mathbf{X}$ contains mean-zero columns and $\mathbf{S} = N^{-1}\mathbf{X}^T\mathbf{X}$. Let $\widetilde{\mathbf{F}} = \mathbf{F}^{(d)} + \widetilde{\mathbf{F}}^{(u)}$ be alternative factors with a different indeterminate part. The following statements hold:*

(i) $\mathbf{F}^{(d)}$ *is the regression of* $\mathbf{F}$ *on* $\mathbf{X}$, *i.e.,* $\mathbf{F}^{(d)} = \mathbf{X}\mathbf{S}^{-1}(N^{-1}\mathbf{X}^T\mathbf{F})$,
(ii) $\mathrm{ssq}(\mathbf{F} - \widetilde{\mathbf{F}})$ *is maximal when* $\widetilde{\mathbf{F}}^{(u)} = -\mathbf{F}^{(u)}$,
(iii) $\mathrm{Corr}(\mathbf{f}_r, \widetilde{\mathbf{f}}_r) \geq 2\,\mathrm{Var}(\mathbf{f}_r^{(d)}) - 1$, *with equality when* $\widetilde{\mathbf{F}}^{(u)} = -\mathbf{F}^{(u)}$, *where* $\mathbf{f}_r^{(d)}$ *is the* r*th column of* $\mathbf{F}^{(d)}$,
(iv) $\mathrm{Var}(\mathbf{F}^{(u)}) = \mathrm{Var}(\mathbf{F}) - \mathrm{Var}(\mathbf{F}^{(d)}) = \mathbf{\Phi} - (N^{-1}\mathbf{X}^T\mathbf{F})^T\mathbf{S}^{-1}(N^{-1}\mathbf{X}^T\mathbf{F})$.

Proof. First, we prove (i). Since $\mathbf{F}^{(d)} \in \mathrm{span}(\mathbf{X})$, we write $\mathbf{F}^{(d)} = \mathbf{X}\mathbf{B}_f$ for some $\mathbf{B}_f \in \mathbb{R}^{J\times R}$. The regression of $\mathbf{F}$ on $\mathbf{X}$ is given by $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{F}$, which equals $\mathbf{X}\mathbf{S}^{-1}(N^{-1}\mathbf{X}^T\mathbf{F})$. Since $\mathbf{X}^T\mathbf{F}^{(u)} = \mathbf{O}$, the regression is equal to $\mathbf{X}\mathbf{B}_f = \mathbf{F}^{(d)}$, which proves (i).

The proofs of (ii) and (iii) are analogous to Guttman (1955). Since the factors have variance one, we have

$$\begin{aligned}\mathrm{ssq}(\mathbf{F} - \widetilde{\mathbf{F}}) &= \mathrm{trace}(\mathbf{F}^T\mathbf{F} - \mathbf{F}^T\widetilde{\mathbf{F}} - \widetilde{\mathbf{F}}^T\mathbf{F} + \widetilde{\mathbf{F}}^T\widetilde{\mathbf{F}}) \\ &= 2RN - 2\,\mathrm{trace}(\mathbf{F}^T\widetilde{\mathbf{F}}) \\ &= 2RN - 2N\,\mathrm{trace}(\mathrm{Corr}(\mathbf{F}, \widetilde{\mathbf{F}})) \\ &= 2RN - 2N\,\mathrm{trace}(N^{-1}(\mathbf{F}^{(d)})^T\mathbf{F}^{(d)} + N^{-1}(\mathbf{F}^{(u)})^T\widetilde{\mathbf{F}}^{(u)}).\end{aligned} \tag{9}$$

Hence, we need to minimize $\mathrm{trace}(N^{-1}(\mathbf{F}^{(u)})^T\widetilde{\mathbf{F}}^{(u)})$. Due to orthogonality of the determinate and indeterminate parts, we have

$$N^{-1}(\mathbf{F}^{(u)})^T\mathbf{F}^{(u)} = N^{-1}(\widetilde{\mathbf{F}}^{(u)})^T\widetilde{\mathbf{F}}^{(u)} = \mathbf{\Phi} - N^{-1}(\mathbf{F}^{(d)})^T\mathbf{F}^{(d)}. \tag{10}$$

Hence, the columns of $\mathbf{F}^{(u)}$ and those of $\widetilde{\mathbf{F}}^{(u)}$ have identical sum of squares. Using the Cauchy–Schwarz inequality, it follows that each diagonal element of $N^{-1}(\mathbf{F}^{(u)})^T\widetilde{\mathbf{F}}^{(u)}$ is minimal for $\widetilde{\mathbf{F}}^{(u)} = -\mathbf{F}^{(u)}$. This proves (ii).

The proof of (iii) follows from (9) and (10) by observing that each diagonal element of $\mathrm{Corr}(\mathbf{F}, \widetilde{\mathbf{F}})$ is minimal for $\widetilde{\mathbf{F}}^{(u)} = -\mathbf{F}^{(u)}$. Moreover, the minimal rth correlation can be written as $\mathrm{Var}(\mathbf{f}_r^{(d)}) - (1 - \mathrm{Var}(\mathbf{f}_r^{(d)})) = 2\,\mathrm{Var}(\mathbf{f}_r^{(d)}) - 1$. Note that the diagonal of $N^{-1}(\mathbf{F}^{(d)})^T\mathbf{F}^{(d)}$ contains $\mathrm{Var}(\mathbf{f}_r^{(d)})$, $r = 1, \ldots, R$, since $\mathbf{F}^{(d)} \in \mathrm{span}(\mathbf{X})$ has mean-zero columns. This completes the proof of (iii).

The proof of (iv) follows from (10) since $\mathrm{Var}(\mathbf{F}^{(d)}) = (N^{-1}\mathbf{X}^T\mathbf{F})^T\mathbf{S}^{-1}(N^{-1}\mathbf{X}^T\mathbf{F})$ by (i). □

As we will see in Section 4, under models I and II we have $\text{Corr}(\mathbf{X}, \widehat{\mathbf{F}}) = N^{-1}\mathbf{X}^T\widehat{\mathbf{F}} = \widehat{\mathbf{P}}\,\widehat{\mathbf{\Phi}}$. Hence, by Lemma 3.1(i) the estimate of the determinate part $\widehat{\mathbf{F}}^{(d)}$ is identical to the least squares factor score estimate of Thurstone (1935). And, since $\text{Var}(\hat{\mathbf{f}}_r^{(d)}) = ((\widehat{\mathbf{P}}\,\widehat{\mathbf{\Phi}})^T\mathbf{S}^{-1}(\widehat{\mathbf{P}}\,\widehat{\mathbf{\Phi}}))_{rr}$, the expression for the minimal correlation in Lemma 3.1(iii) is analogous to the case of perfect fit in (8).

Note that $\text{Var}(\widehat{\mathbf{F}}^{(d)})$ expresses the sampling uncertainty of the least squares estimate of the factor scores, while $\text{Var}(\widehat{\mathbf{F}}^{(u)})$ expresses the uncertainty in the factor scores due to factor indeterminacy. We have $\text{Var}(\hat{\mathbf{f}}_r^{(d)}) = 1 - \text{Var}(\hat{\mathbf{f}}_r^{(u)})$. The same expressions for $\text{Var}(\widehat{\mathbf{F}}^{(d)})$ and $\text{Var}(\widehat{\mathbf{F}}^{(u)})$ as above can be derived within the empirical Bayes framework, with $\text{Var}(\widehat{\mathbf{F}}^{(u)})$ being the posterior variance of $\widehat{\mathbf{F}}$ given the data $\mathbf{X}$ and $\widehat{\mathbf{P}}$, $\widehat{\mathbf{U}}$, and $\widehat{\mathbf{\Phi}}$ (Skrondal and Rabe-Hesketh, 2004).

Finally, we consider the empirical finding of Schönemann and Wang (1972) and Grice (2001) that weak factors (with less explained variance) tend to have smaller minimal correlations. In the general setting of Lemma 3.1, we give a mathematical explanation for this, which is a novel result. For ease of presentation we consider the case of orthogonal factors ($\mathbf{\Phi} = \mathbf{I}_R$). Let $\mathbf{F}^{(d)} \in \text{span}(\mathbf{X})$ be given by $\mathbf{F}^{(d)} = \mathbf{X}\,\mathbf{B}_f$ for some $\mathbf{B}_f \in \mathbb{R}^{J\times R}$. We have

$$\text{Var}(\mathbf{F}^{(d)}) = N^{-1}(\mathbf{F}^{(d)})^T\mathbf{F}^{(d)} = \mathbf{B}_f^T\mathbf{S}\,\mathbf{B}_f. \tag{11}$$

The explained common variances are found on the diagonal of $\mathbf{P}\mathbf{P}^T$. Suppose $\mathbf{P} = N^{-1}\mathbf{X}^T\mathbf{F}$, which is equal to $\mathbf{P} = N^{-1}\mathbf{X}^T\mathbf{F}^{(d)} = \mathbf{S}\,\mathbf{B}_f$. This is a common assumption in factor analysis and holds under models I and II; see Section 4. The explained common variances due to each factor are on the diagonal of

$$\mathbf{P}^T\mathbf{P} = \mathbf{B}_f^T\mathbf{S}^2\,\mathbf{B}_f. \tag{12}$$

As we see, the expressions in (11) and (12) are closely related. Let $\mathbf{S} = \mathbf{Q}\,\mathbf{\Lambda}\,\mathbf{Q}^T$ be the eigen decomposition of $\mathbf{S}$, with $\mathbf{Q} \in \mathbb{R}^{J\times J}$ such that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_J$, and diagonal positive semi-definite $\mathbf{\Lambda} \in \mathbb{R}^{J\times J}$. Let $\mathbf{C} = \mathbf{Q}^T\mathbf{B}_f$. For the rth factor we obtain

$$(\mathbf{B}_f^T\mathbf{S}\,\mathbf{B}_f)_{rr} = \sum_{j=1}^{J} c_{jr}^2\,\lambda_j, \qquad (\mathbf{B}_f^T\mathbf{S}^2\,\mathbf{B}_f)_{rr} = \sum_{j=1}^{J} c_{jr}^2\,\lambda_j^2. \tag{13}$$

Hence, if the rth factor has low explained common variance, then also the variance of its determinate part is likely to be low. The latter results in a low minimal correlation and, hence, worse factor indeterminacy. This is a mathematical explanation of the phenomenon observed in Schönemann and Wang (1972) and Grice (2001).

4. The data factor models and simultaneous estimation algorithms

In Section 4.1 we introduce the data factor models I and II formally. The simultaneous estimation procedures for models I and II are given in Sections 4.2 and 4.3, respectively. In Section 4.4 we show how the expression of the indeterminate part of $\widehat{\mathbf{F}}$ can be used to simulate probability density functions for the factor scores due to indeterminacy. Additional technical remarks on the data factor models can be found in the Appendix.

4.1. Data factor models I and II

In practice, the data factor model (4) does not fit perfectly and there is a residual. We consider the following two options for the residual:

$$\text{Data factor model I:} \quad \mathbf{X} = \overbrace{\mathbf{F}\,\mathbf{P}^T}^{\text{common part}} + \overbrace{\mathbf{E}\,\mathbf{U}}^{\text{unique part}} + \text{residual}$$

$$\text{Data factor model II:} \quad \mathbf{X} = (\mathbf{F}\,\mathbf{P}^T + \text{residual}) + \mathbf{E}\,\mathbf{U}$$

Both models are fitted by minimizing the sum of squares of the residuals under the following assumptions:

$$\text{Data factor model I}: \min_{\substack{N^{-1}[\mathbf{F}\ \mathbf{E}]^T[\mathbf{F}\ \mathbf{E}]=\mathbf{I}_{R+J}\\ \mathbf{F}^T\mathbf{1}_N=\mathbf{0},\ \mathbf{E}^T\mathbf{1}_N=\mathbf{0}\\ \mathbf{U}\ \text{diagonal}}} \text{ssq}(\mathbf{X} - \mathbf{F}\,\mathbf{P}^T - \mathbf{E}\,\mathbf{U}), \tag{14}$$

$$\text{Data factor model II}: \min_{\substack{N^{-1}[\mathbf{F}\ \mathbf{E}]^T[\mathbf{F}\ \mathbf{E}]=\mathbf{I}_{R+J}\\ \mathbf{F}^T\mathbf{1}_N=\mathbf{0},\ \mathbf{E}^T\mathbf{1}_N=\mathbf{0}\\ \mathbf{U}\ \text{diagonal}\\ N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{E}\mathbf{U})=\mathbf{O}}} \text{ssq}(\mathbf{X} - \mathbf{F}\,\mathbf{P}^T - \mathbf{E}\,\mathbf{U}). \tag{15}$$

The relations between the common and unique parts and the residuals may deserve some explanation. Recall that the unique part of an observed variable represents either measurement error or a specific part measured only by that variable, and is uncorrelated to the common part and the unique parts of other variables. In model I the common part is fully explained

by the common factors, while in model II the common factors explain only part of the common part. Under model I the residuals are not part of the common or unique parts and the residual sum of squares can be interpreted as unexplained variance in the observed variables. Under model II the residuals are that part of the common part that is not explained by the common factors and the residual sum of squares can be interpreted as the unexplained common variance (see Theorem 4.2 below). Under both models the common and unique parts are uncorrelated, in model II due to the additional constraint $N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{EU})=\mathbf{O}$.

In both models $\mathbf{F}$ and $\mathbf{E}$ have mean-zero columns, and are uncorrelated to each other and within themselves. Although models I and II assume and estimate orthogonal factors, the estimated factors and loadings can be rotated orthogonally or obliquely without loss of fit. That is, $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{P}}$ can be replaced by rotated versions $\widehat{\mathbf{F}}\mathbf{T}$ and $\widehat{\mathbf{P}}\mathbf{T}^{-T}$ for a nonsingular $R\times R$ matrix $\mathbf{T}$ with column sum of squares equal to one. The latter guarantees that the factor correlation matrix after rotation $\widehat{\mathbf{\Phi}}=\mathbf{T}^T\mathbf{T}$ has a diagonal of ones. As we will see in Sections 4.2 and 4.3, we have $\widehat{\mathbf{P}}=N^{-1}\mathbf{X}^T\widehat{\mathbf{F}}$. After rotating we thus obtain $(\widehat{\mathbf{P}}\mathbf{T}^{-T})\widehat{\mathbf{\Phi}}=N^{-1}\mathbf{X}^T(\widehat{\mathbf{F}}\mathbf{T})$.

Model I was introduced independently by Kiers in Sočan (2003) and by De Leeuw (2004). Both derived the same alternating least squares (ALS) algorithm to update each of $[\mathbf{F}\ \ \mathbf{E}]$, $\mathbf{P}$, and $\mathbf{U}$ while keeping the others fixed. De Leeuw (2004, 2008) also shows that the estimate $[\widehat{\mathbf{F}}\ \ \widehat{\mathbf{E}}]$ can be written as the sum of a determinate part and an indeterminate part, analogous to the analysis of Guttman (1955) for perfect fit. Model I has been adapted to the case of more variables than observations (i.e., when $J>N$) by Unkel and Trendafilov (2010, 2013) and Trendafilov and Unkel (2011). Unkel and Trendafilov (2010) also contains a nice overview of simultaneous estimation methods in common factor analysis.

We introduce the novel data factor model II and derive a simultaneous estimation algorithm, and decompositions of $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{E}}$ into determinate and indeterminate parts. It will be shown that the estimation algorithm is not iterative as ALS, but first estimates $\mathbf{U}^2$ by MRFA, then computes $\widehat{\mathbf{E}}$, and finally computes $\widehat{\mathbf{F}}\widehat{\mathbf{P}}^T$ as a best rank-R approximation of $\mathbf{X}-\widehat{\mathbf{E}}\widehat{\mathbf{U}}$. In fact, model II is the simultaneous estimation analogue of MRFA and yields the MRFA estimates of $\mathbf{P}$ and $\mathbf{U}$. In this sense, MRFA and model II have the same relation as MLFA and maximum likelihood ratio estimation of the fixed factor model as shown by McDonald (1979). The advantage of model II over model I is the same as the advantage of MRFA over MINRES or MLFA: we are able to compute the percentage of explained common variance, which can be used to choose the number of factors. Large sample theory for MRFA under normality can be found in Shapiro and Ten Berge (2002). In particular, the asymptotic bias and its variance of the estimated explained common variance are derived, as well as the asymptotic covariance matrix of the estimated unique variances. These large sample results apply to the data factor model II as well. Although no asymptotic standard errors for the estimated loadings are available, finite sample standard errors can be obtained via a bootstrap procedure (e.g., Zhang, 2014) as will be demonstrated in the application in Section 6.

4.2. Algorithm for model I

In the ALS algorithm derived by Kiers in Sočan (2003) and by De Leeuw (2004), each of $[\mathbf{F}\ \mathbf{E}]$, $\mathbf{P}$, and $\mathbf{U}$ is updated while keeping the others fixed. First, we describe the update of $[\mathbf{F}\ \ \mathbf{E}]$, for fixed $\mathbf{P}$ and $\mathbf{U}$. The objective function in (14) can be written as

$$\operatorname{ssq}(\mathbf{X}-[\mathbf{F}\,\mathbf{E}]\,[\mathbf{P}\,\mathbf{U}]^T)=-2\operatorname{trace}([\mathbf{F}\,\mathbf{E}]^T\mathbf{X}\,[\mathbf{P}\,\mathbf{U}])+(\text{const}),\tag{16}$$

where $(\text{const})=\operatorname{trace}(\mathbf{X}^T\mathbf{X})+N\operatorname{trace}([\mathbf{P}\ \mathbf{U}]\,[\mathbf{P}\ \mathbf{U}]^T)$ does not depend on $\mathbf{F}$ or $\mathbf{E}$. De Leeuw (2004) proves that all optimal $[\mathbf{F}\ \mathbf{E}]$ are given by

$$[\mathbf{F}\ \mathbf{E}]=N^{1/2}\,\mathbf{Q}_1\,\mathbf{Q}_2^T,\tag{17}$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are obtained from the singular value decomposition $\mathbf{X}\,[\mathbf{P}\ \ \mathbf{U}]=\mathbf{Q}_1\,\mathbf{D}\,\mathbf{Q}_2^T$, with $\mathbf{Q}_1\in\mathbb{R}^{N\times(J+R)}$ having mean-zero orthonormal columns, $\mathbf{Q}_2\in\mathbb{R}^{(J+R)\times(J+R)}$ having orthonormal columns, and $\mathbf{D}\in\mathbb{R}^{(J+R)\times(J+R)}$ the diagonal matrix containing the singular values in decreasing order. Note that $\operatorname{rank}(\mathbf{X}\,[\mathbf{P}\ \mathbf{U}])\leq J$ implies that the last R singular values in $\mathbf{D}$ are zero and, hence, that the last R columns of $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are indeterminate. We write $\mathbf{Q}_1=[\mathbf{Q}_1^{(d)}\ \mathbf{Q}_1^{(u)}]$ and $\mathbf{Q}_2=[\mathbf{Q}_2^{(d)}\ \mathbf{Q}_2^{(u)}]$, with superscript (d) denoting the J determinate columns, and (u) denoting the R indeterminate columns. We rewrite (17) as $[\mathbf{F}\ \mathbf{E}]=N^{1/2}\,\mathbf{Q}_1^{(d)}\,(\mathbf{Q}_2^{(d)})^T+N^{1/2}\,\mathbf{Q}_1^{(u)}\,(\mathbf{Q}_2^{(u)})^T$. Since $\mathbf{Q}_1^{(d)}\in\operatorname{span}(\mathbf{X})$ and $\mathbf{Q}_1^{(u)}\in\operatorname{span}(\mathbf{X})^{\perp}$, it follows that both $\mathbf{F}$ and $\mathbf{E}$ are decomposed into a determinate part in $\operatorname{span}(\mathbf{X})$ and an indeterminate part in $\operatorname{span}(\mathbf{X})^{\perp}$. Hence, the results of Lemma 3.1 apply to the update of $\mathbf{F}$ under model I.

The updates of $\mathbf{P}$ and $\mathbf{U}$ are given as $\mathbf{P}=N^{-1}\mathbf{X}^T\mathbf{F}$ and $\mathbf{U}=N^{-1}\operatorname{diag}(\mathbf{E}^T\mathbf{X})$ (Sočan, 2003; De Leeuw, 2004). Note that these are not influenced by the indeterminate parts of $\mathbf{F}$ and $\mathbf{E}$, which are orthogonal to $\mathbf{X}$.

For later use, we consider deriving the update of $[\mathbf{F}\ \mathbf{E}]$ under model I with Lagrange multipliers. Let $\mathbf{T}=N^{-1/2}\,[\mathbf{F}\ \mathbf{E}]$ and $\mathbf{Y}=N^{1/2}\,\mathbf{X}\,[\mathbf{P}\ \mathbf{U}]$. By (16), we maximize $g(\mathbf{T})=\operatorname{trace}(\mathbf{T}^T\,\mathbf{Y})$ subject to $\mathbf{T}^T\mathbf{T}=\mathbf{I}_{J+R}$. The mean-zero constraint on $\mathbf{T}$ is ignored. We augment $g(\mathbf{T})$ with a term containing Lagrange multipliers corresponding to $\mathbf{T}^T\mathbf{T}=\mathbf{I}_{J+R}$, and obtain

$$h_{\mathrm{I}}(\mathbf{T},\mathbf{L})=\operatorname{trace}(\mathbf{T}^T\,\mathbf{Y})+\operatorname{trace}(\mathbf{L}\,(\mathbf{T}^T\mathbf{T}-\mathbf{I}_{J+R})),\tag{18}$$

where $\mathbf{L}\in\mathbb{R}^{(J+R)\times(J+R)}$ is a matrix of Lagrange multipliers. The update of $\mathbf{T}$ is found by solving $\partial h_{\mathrm{I}}/\partial\mathbf{T}=\mathbf{O}$ and $\partial h_{\mathrm{I}}/\partial\mathbf{L}=\mathbf{O}$, where the latter is equivalent to $\mathbf{T}^T\mathbf{T}=\mathbf{I}_{J+R}$.

4.3. Algorithm for model II

Before we derive the algorithm for fitting model II, we establish some preliminary results that we need later. First, we prove that the estimates of $\mathbf{F}$ and $\mathbf{E}$ can be decomposed into determinate and indeterminate parts as under model I above. We use the method of Lagrange multipliers for this. Under model II, the additional constraint $N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{EU})=\mathbf{O}$ can be written as $N^{-1}\mathbf{E}^T\mathbf{X}=\mathbf{U}$. Substituting into the objective function implies the maximization of trace$(\mathbf{F}^T\mathbf{X}\mathbf{P})$ analogous to (16). Let $\mathbf{T}_1=N^{-1/2}\mathbf{F}$, $\mathbf{T}_2=N^{-1/2}\mathbf{E}$, and $\mathbf{T}=[\mathbf{T}_1\ \mathbf{T}_2]$. We ignore the mean-zero constraints on $\mathbf{T}_1$ and $\mathbf{T}_2$. The objective function including Lagrange multipliers is given by

$$\begin{aligned} h_{\text{II}}(\mathbf{T}_1,\mathbf{T}_2,\mathbf{K},\mathbf{L}) &= \text{trace}(\mathbf{T}_1^T N^{1/2}\mathbf{X}\mathbf{P})+\text{trace}(\mathbf{L}(\mathbf{T}^T\mathbf{T}-\mathbf{I}_{J+R}))+\text{trace}(\mathbf{K}(N^{1/2}\mathbf{T}_2^T\mathbf{X}-N\mathbf{U})) \\ &= \text{trace}(\mathbf{T}^T N^{1/2}\mathbf{X}[\mathbf{P}\ \mathbf{K}])+\text{trace}(\mathbf{L}(\mathbf{T}^T\mathbf{T}-\mathbf{I}_{J+R}))-\text{trace}(N\mathbf{U}\mathbf{K}), \end{aligned} \tag{19}$$

where $\mathbf{L}\in\mathbb{R}^{(J+R)\times(J+R)}$ and $\mathbf{K}\in\mathbb{R}^{J\times J}$ contain Lagrange multipliers. Comparing (19) to (18), we see that we obtain the same derivative form $\partial h/\partial\mathbf{T}$ as under model I, with $\mathbf{X}[\mathbf{P}\ \mathbf{K}]$ replacing $\mathbf{X}[\mathbf{P}\ \mathbf{U}]$. It follows that, for each $\mathbf{K}$, all solutions are given by $[\mathbf{F}\ \mathbf{E}]=N^{1/2}\mathbf{Q}_1\mathbf{Q}_2^T$, where the singular value decomposition of $\mathbf{X}[\mathbf{P}\ \mathbf{K}]$ is given by $\mathbf{Q}_1\mathbf{D}\mathbf{Q}_2^T$. As under model I, the estimate of $[\mathbf{F}\ \mathbf{E}]$ is decomposed into a determinate part in span$(\mathbf{X})$ and an indeterminate part in span$(\mathbf{X})^\perp$. Hence, the results of Lemma 3.1 also apply to the estimate of $\mathbf{F}$ under model II.

Next, we show that under model II, $\mathbf{S}-\mathbf{U}^2$ is positive semi-definite (denoted as ≥ 0). This is analogous to the MRFA method of Ten Berge and Kiers (1991).

Lemma 4.1. *Under model* II *in* (15)*, we have* $\mathbf{S}-\mathbf{U}^2\geq 0$.

Proof. Using the orthogonality between $\mathbf{E}$ and $\mathbf{X}-\mathbf{EU}$, we obtain

$$\begin{aligned} N^{-1}(\mathbf{X}-\mathbf{EU})^T(\mathbf{X}-\mathbf{EU}) &= N^{-1}\mathbf{X}^T\mathbf{X}-N^{-1}\mathbf{X}^T(\mathbf{EU}) \\ &= N^{-1}\mathbf{X}^T\mathbf{X}-N^{-1}(\mathbf{X}-\mathbf{EU})^T(\mathbf{EU})-N^{-1}(\mathbf{EU})^T(\mathbf{EU}) \\ &= N^{-1}\mathbf{X}^T\mathbf{X}-N^{-1}(\mathbf{EU})^T(\mathbf{EU}) \\ &= \mathbf{S}-\mathbf{U}^2\geq 0, \end{aligned} \tag{20}$$

where the positive semi-definiteness follows from equality with the left-hand side. □

The common part $\mathbf{X}-\mathbf{EU}$ is approximated by $\mathbf{FP}^T$. It follows from (20) that the percentage of explained common variance under model II is the same as in MRFA, namely

$$100\ \cdot\ \frac{\text{trace}(\mathbf{P}\boldsymbol{\Phi}\mathbf{P}^T)}{\text{trace}(\mathbf{S}-\mathbf{U}^2)}. \tag{21}$$

For variable j, the explained common variance can be computed as $100\cdot(\mathbf{P}\boldsymbol{\Phi}\mathbf{P}^T)_{jj}/(1-u_j^2)$. For orthogonal factors, the explained common variance due to each factor is computed as $100\cdot(\mathbf{P}^T\mathbf{P})_{rr}/\text{trace}(\mathbf{S}-\mathbf{U}^2)$. Under model I it is not meaningful to compute the explained common variance, since the unique parts are not required to be orthogonal to $\mathbf{X}-\mathbf{EU}$ and, hence, may contain some unexplained common variance.

Now we are ready to derive the algorithm for model II. Contrary to the ALS algorithm for model I, the algorithm for model II is not iterative. First, we present a result on the optimal $\mathbf{F}$ and $\mathbf{P}$ for given $\mathbf{E}$ and $\mathbf{U}$. Let the singular value decomposition of $\mathbf{X}-\mathbf{EU}$ be given by $\mathbf{Q}_3(N^{1/2}\boldsymbol{\Psi})\mathbf{Q}_4^T$, with $\mathbf{Q}_3\in\mathbb{R}^{N\times J}$ and $\mathbf{Q}_4\in\mathbb{R}^{J\times J}$ such that $\mathbf{Q}_3^T\mathbf{Q}_3=\mathbf{Q}_4^T\mathbf{Q}_4=\mathbf{I}_J$, and diagonal positive semi-definite $\boldsymbol{\Psi}\in\mathbb{R}^{J\times J}$ with diagonal entries in decreasing order. This implies $\mathbf{S}-\mathbf{U}^2=\mathbf{Q}_4\boldsymbol{\Psi}^2\mathbf{Q}_4^T$. Let $\mathbf{Q}_{3,R}\in\mathbb{R}^{N\times R}$ and $\mathbf{Q}_{4,R}\in\mathbb{R}^{J\times R}$ contain the first R columns of $\mathbf{Q}_3$ and $\mathbf{Q}_4$, respectively, and let $\boldsymbol{\Psi}_R\in\mathbb{R}^{R\times R}$ be diagonal and contain the first R diagonal entries of $\boldsymbol{\Psi}$.

Theorem 4.2. *Let* $\mathbf{E}$ *and* $\mathbf{U}$ *be given such that* $N^{-1}\mathbf{E}^T\mathbf{E}=\mathbf{I}_J$, $\mathbf{E}^T\mathbf{1}_N=\mathbf{0}$, *and* $N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{EU})=\mathbf{O}$. *Let* $\psi_R\neq\psi_{R+1}$. *Under model II, all optimal* $\mathbf{F}$ *and* $\mathbf{P}$ *(up to orthogonal factor rotation) are given by*

$$\mathbf{F}=N^{1/2}\mathbf{Q}_{3,R}=(\mathbf{X}-\mathbf{EU})\mathbf{Q}_{4,R}\boldsymbol{\Psi}_R^{-1},\qquad \mathbf{P}=\mathbf{Q}_{4,R}\boldsymbol{\Psi}_R. \tag{22}$$

The minimal objective function equals

$$\text{ssq}(\mathbf{X}-\mathbf{F}\mathbf{P}^T-\mathbf{E}\mathbf{U})=N(\psi_{R+1}^2+\cdots+\psi_J^2), \tag{23}$$

where $\psi_{R+1}^2,\ldots,\psi_J^2$ *are the* $J-R$ *smallest singular values of* $\mathbf{S}-\mathbf{U}^2$.

Proof. Taking $\mathbf{F}$ and $\mathbf{P}$ as in (22) implies that $\mathbf{FP}^T$ is a best rank-R approximation of the common part $\mathbf{X}-\mathbf{EU}$ (Eckart and Young, 1936). The minimizer $\mathbf{FP}^T$ is unique if and only if $\psi_R\neq\psi_{R+1}$. The minimal approximation error is given by $\text{ssq}(\mathbf{X}-\mathbf{F}\mathbf{P}^T-\mathbf{E}\mathbf{U})=N(\psi_{R+1}^2+\cdots+\psi_J^2)$ (Eckart and Young, 1936). When $\mathbf{F}$ is chosen as in (22), it follows that $N^{-1}\mathbf{F}^T\mathbf{F}=\mathbf{I}_R$, that $\mathbf{E}^T\mathbf{F}=\mathbf{O}$ (since $N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{EU})=\mathbf{O}$), and that $\mathbf{F}$ has mean-zero columns (since $\mathbf{X}-\mathbf{EU}$ has mean-zero columns). Hence, $\mathbf{F}$ satisfies the assumptions of the model. □

Note that the minimal unexplained common variance depends on the singular values of $\mathbf{S}-\mathbf{U}^2$, but not on the unique part $\mathbf{E}$. Hence, for any $\mathbf{E}$ satisfying the assumptions of the model, choosing $\mathbf{F}$ and $\mathbf{P}$ as in (22) is optimal.

Next, we show that for any $\mathbf{U}$ such that $\mathbf{S}-\mathbf{U}^2 \geq 0$ we can find an $\mathbf{E}$ satisfying the assumptions of the model.

Theorem 4.3. *Let* $\mathbf{U}$ *be such that* $\mathbf{S}-\mathbf{U}^2 \geq 0$. *Then we can find an* $\mathbf{E}$ *satisfying the assumptions of model II, that is* $N^{-1}\mathbf{E}^T\mathbf{E}=\mathbf{I}_J$, $\mathbf{E}^T\mathbf{1}_N=\mathbf{0}$, *and* $N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{E}\mathbf{U})=\mathbf{O}$.

Proof. From the above, we know that $\mathbf{E}=\mathbf{E}^{(d)}+\mathbf{E}^{(u)}$, where the determinate part $\mathbf{E}^{(d)} \in \text{span}(\mathbf{X})$ and the indeterminate part $\mathbf{E}^{(u)} \in \text{span}(\mathbf{X}, \mathbf{1}_N)^\perp$. The $\mathbf{1}_N$ in the latter span ensures that $\mathbf{E}$ has mean-zero columns. By using $N^{-1}\mathbf{E}^T\mathbf{E}=\mathbf{I}_J$, the constraint $N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{E}\mathbf{U})=\mathbf{O}$ can be written as $N^{-1}\mathbf{E}^T\mathbf{X}=\mathbf{U}$, which is identical to $N^{-1}(\mathbf{E}^{(d)})^T\mathbf{X}=\mathbf{U}$. Since $\mathbf{E}^{(d)} \in \text{span}(\mathbf{X})$, we write $\mathbf{E}^{(d)}=\mathbf{X}\mathbf{B}_e$ for some $\mathbf{B}_e \in \mathbb{R}^{J\times J}$. The constraint becomes $N^{-1}\mathbf{B}_e^T\mathbf{X}^T\mathbf{X}=\mathbf{B}_e^T\mathbf{S}=\mathbf{U}$, which implies

$$\mathbf{E}^{(d)}=\mathbf{X}\mathbf{S}^{-1}\mathbf{U}, \tag{24}$$

where we assume that $\mathbf{S}$ is nonsingular. Hence, the columns of $\mathbf{E}^{(d)}$ are determined by the data $\mathbf{X}$ up to their magnitude. Note that this expression for $\mathbf{E}^{(d)}$ is identical to the case of perfect fit (6). Using (24), the constraint $N^{-1}\mathbf{E}^T\mathbf{E}=\mathbf{I}_J$ can be written as

$$N^{-1}(\mathbf{E}^{(u)})^T\mathbf{E}^{(u)}=\mathbf{I}_J-N^{-1}(\mathbf{E}^{(d)})^T\mathbf{E}^{(d)}=\mathbf{I}_J-N^{-1}\mathbf{U}\mathbf{S}^{-1}\mathbf{X}^T\mathbf{X}\mathbf{S}^{-1}\mathbf{U}=\mathbf{I}_J-\mathbf{U}\mathbf{S}^{-1}\mathbf{U}. \tag{25}$$

Hence, the matrix $\mathbf{I}_J-\mathbf{U}\mathbf{S}^{-1}\mathbf{U}$ must be positive semi-definite. By Proposition 4.4 below, this is equivalent to $\mathbf{S}-\mathbf{U}^2$ being positive semi-definite.

The indeterminate part $\mathbf{E}^{(u)}$ can be computed as follows. The indeterminate part $\mathbf{E}^{(u)}$ satisfies (25) and lies in $\text{span}(\mathbf{X}, \mathbf{1}_N)^\perp$. Let $\mathbf{G} \in \mathbb{R}^{N\times(N-J-1)}$ contain as columns an orthonormal basis of $\text{span}(\mathbf{X}, \mathbf{1}_N)^\perp$. We write $\mathbf{E}^{(u)}=\mathbf{G}\mathbf{H}$ for some $\mathbf{H} \in \mathbb{R}^{(N-J-1)\times J}$. Let the singular value decomposition of $\mathbf{I}_J-\mathbf{U}\mathbf{S}^{-1}\mathbf{U} \geq 0$ be given by $\mathbf{Q}_5\,\mathbf{\Gamma}\,\mathbf{Q}_5^T$, with $\mathbf{Q}_5 \in \mathbb{R}^{J\times\tilde{R}}$ such that $\mathbf{Q}_5^T\mathbf{Q}_5=\mathbf{I}_{\tilde{R}}$, diagonal positive semi-definite $\mathbf{\Gamma} \in \mathbb{R}^{\tilde{R}\times\tilde{R}}$, and $\tilde{R}=\text{rank}(\mathbf{I}_J-\mathbf{U}\mathbf{S}^{-1}\mathbf{U})$. By (25), it follows that $N^{-1}\mathbf{E}^T\mathbf{E}=\mathbf{I}_J$ is satisfied when $\mathbf{H}^T\mathbf{H}=\mathbf{Q}_5\,(N\,\mathbf{\Gamma})\,\mathbf{Q}_5^T$, where we used $\mathbf{G}^T\mathbf{G}=\mathbf{I}_{N-J-1}$. Hence, we obtain that

$$\mathbf{E}^{(u)}=N^{1/2}\,\mathbf{G}\,\mathbf{M}\,\mathbf{\Gamma}^{1/2}\,\mathbf{Q}_5^T, \tag{26}$$

where $\mathbf{M} \in \mathbb{R}^{(N-J-1)\times\tilde{R}}$ is arbitrary such that $\mathbf{M}^T\mathbf{M}=\mathbf{I}_{\tilde{R}}$. This shows explicitly the arbitrariness in the indeterminate part $\mathbf{E}^{(u)}$.

To sum up, we have $\mathbf{E}^T\mathbf{1}_N=\mathbf{0}$ since both $\mathbf{E}^{(d)}$ in (24) and $\mathbf{E}^{(u)}$ in (26) have mean-zero columns. And for any $\mathbf{U}$ such that $\mathbf{S}-\mathbf{U}^2 \geq 0$, we have $N^{-1}\mathbf{E}^T\mathbf{E}=\mathbf{I}_J$ and $N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{E}\mathbf{U})=\mathbf{O}$ for $\mathbf{E}^{(d)}$ in (24) and $\mathbf{E}^{(u)}$ in (26). This completes the proof. □

Proposition 4.4. *Let* $\mathbf{S}$ *be* $J\times J$ *and positive definite and* $\mathbf{U}$ *a* $J\times J$ *diagonal matrix. Then* $\mathbf{I}_J-\mathbf{U}\mathbf{S}^{-1}\mathbf{U}$ *is positive semi-definite if and only if* $\mathbf{S}-\mathbf{U}^2$ *is positive semi-definite.*

Proof. Pre- and postmultiplying $\mathbf{S}-\mathbf{U}^2$ by $\mathbf{S}^{-1/2}$ does not affect its positive semi-definiteness. Hence, $\mathbf{S}-\mathbf{U}^2 \geq 0$ if and only if $\mathbf{I}_J-\mathbf{S}^{-1/2}\mathbf{U}^2\mathbf{S}^{-1/2} \geq 0$. The eigenvalues of $\mathbf{S}^{-1/2}\mathbf{U}^2\mathbf{S}^{-1/2}=(\mathbf{U}\mathbf{S}^{-1/2})^T(\mathbf{U}\mathbf{S}^{-1/2})$ are identical to those of $(\mathbf{U}\mathbf{S}^{-1/2})(\mathbf{U}\mathbf{S}^{-1/2})^T=\mathbf{U}\mathbf{S}^{-1}\mathbf{U}$. This implies that the eigenvalues of $\mathbf{I}_J-\mathbf{S}^{-1/2}\mathbf{U}^2\mathbf{S}^{-1/2}$ are identical to those of $\mathbf{I}_J-\mathbf{U}\mathbf{S}^{-1}\mathbf{U}$. Hence, we obtain that $\mathbf{I}_J-\mathbf{S}^{-1/2}\mathbf{U}^2\mathbf{S}^{-1/2} \geq 0$ if and only if $\mathbf{I}_J-\mathbf{U}\mathbf{S}^{-1}\mathbf{U} \geq 0$. This completes the proof. □

It follows from Lemma 4.1, Theorems 4.2 and 4.3 that we may first find $\mathbf{U} \geq 0$ such that $\mathbf{S}-\mathbf{U}^2 \geq 0$ and $N\,(\psi_{R+1}^2+\cdots+\psi_J^2)$ in (23) is minimal. This problem in $\mathbf{U}$ is solved by MRFA (Ten Berge and Kiers, 1991), in which $N\,(\psi_{R+1}^2+\cdots+\psi_J^2)$ denotes the unexplained common variance. Hence, we use the MRFA algorithm of Ten Berge and Kiers (1991) to estimate $\mathbf{U}$. After an optimal $\widehat{\mathbf{U}}$ has been found, we compute $\widehat{\mathbf{E}}$ from (24) and (26). Finally, we obtain $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{P}}$ as in (22). As in MRFA, $\widehat{\mathbf{P}}$ is obtained from the truncated singular value decomposition of $\mathbf{S}-\widehat{\mathbf{U}}^2$ (see Theorem 4.2). Hence, the estimates of both $\mathbf{P}$ and $\mathbf{U}$ are the same as in MRFA.

The steps of the algorithm to fit model II in (15) are now as follows:

1. Compute $\widehat{\mathbf{U}}$ by MRFA.
2. Compute $\widehat{\mathbf{E}}^{(d)}$ by (24).
3. Compute $\widehat{\mathbf{E}}^{(u)}$ by (26), where $\mathbf{I}_J-\widehat{\mathbf{U}}\mathbf{S}^{-1}\widehat{\mathbf{U}} \geq 0$ has singular value decomposition $\mathbf{Q}_5\,\mathbf{\Gamma}\,\mathbf{Q}_5^T$, $\mathbf{G} \in \mathbb{R}^{N\times(N-J-1)}$ is an orthonormal basis for $\text{span}(\mathbf{X}, \mathbf{1}_N)^\perp$, and $\mathbf{M} \in \mathbb{R}^{(N-J-1)\times\tilde{R}}$ is arbitrary such that $\mathbf{M}^T\mathbf{M}=\mathbf{I}_{\tilde{R}}$. Set $\widehat{\mathbf{E}}=\widehat{\mathbf{E}}^{(d)}+\widehat{\mathbf{E}}^{(u)}$.
4. Compute $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{P}}$ by (22), where $\mathbf{X}-\widehat{\mathbf{E}}\widehat{\mathbf{U}}$ has singular value decomposition $\mathbf{Q}_3\,(N^{1/2}\mathbf{\Psi})\,\mathbf{Q}_4^T$.

Note that (22) implies $\widehat{\mathbf{P}}=N^{-1}\mathbf{X}^T\widehat{\mathbf{F}}$. Indeed, we have

$$\begin{aligned}
N^{-1}\mathbf{X}^T\widehat{\mathbf{F}} &= N^{-1}\mathbf{X}^T(\mathbf{X}-\widehat{\mathbf{E}}\widehat{\mathbf{U}})\,\mathbf{Q}_{4,R}\,\mathbf{\Psi}_R^{-1}\\
&= N^{-1}(\mathbf{X}-\widehat{\mathbf{E}}\widehat{\mathbf{U}})^T(\mathbf{X}-\widehat{\mathbf{E}}\widehat{\mathbf{U}})\,\mathbf{Q}_{4,R}\,\mathbf{\Psi}_R^{-1}\\
&= \mathbf{Q}_4\,\mathbf{\Psi}^2\,\mathbf{Q}_4^T\,\mathbf{Q}_{4,R}\,\mathbf{\Psi}_R^{-1}\\
&= \mathbf{Q}_{4,R}\,\mathbf{\Psi}_R=\widehat{\mathbf{P}},
\end{aligned} \tag{27}$$

where we used the constraint $N^{-1}\widehat{\mathbf{E}}^T(\mathbf{X}-\widehat{\mathbf{E}}\widehat{\mathbf{U}})=\mathbf{O}$, and the singular value decomposition of $N^{-1}(\mathbf{X}-\widehat{\mathbf{E}}\widehat{\mathbf{U}})^T(\mathbf{X}-\widehat{\mathbf{E}}\widehat{\mathbf{U}})=\mathbf{S}-\widehat{\mathbf{U}}^2$, which equals $\mathbf{Q}_4\,\mathbf{\Psi}^2\,\mathbf{Q}_4^T$.

4.4. Probability densities of factor scores due to indeterminacy

Here, we consider the expressions for the determinate and indeterminate parts of $\widehat{\mathbf{F}}$. We have $\widehat{\mathbf{F}} = \widehat{\mathbf{F}}^{(d)} + \widehat{\mathbf{F}}^{(u)}$, where $\widehat{\mathbf{F}}^{(d)} \in \text{span}(\mathbf{X})$ and $\widehat{\mathbf{F}}^{(u)} \in \text{span}(\mathbf{X})^{\perp}$. These can be used to simulate probability densities of estimated factor scores due to indeterminacy. Under model I, expressions for $\widehat{\mathbf{F}}^{(d)}$ and $\widehat{\mathbf{F}}^{(u)}$ are given by Lemma 3.1 (as shown in Section 4.2). Below, we focus on model II. By Lemma 3.1(i), we have

$$\begin{aligned}
\widehat{\mathbf{F}}^{(d)} &= \mathbf{X}\,(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\widehat{\mathbf{F}} \\
&= \mathbf{X}\,(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\,(\mathbf{X} - \widehat{\mathbf{E}}\widehat{\mathbf{U}})\,\mathbf{Q}_{4,R}\,\boldsymbol{\Psi}_R^{-1} \\
&= (\mathbf{X} - \mathbf{X}\,(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\widehat{\mathbf{E}}^{(d)}\widehat{\mathbf{U}})\,\mathbf{Q}_{4,R}\,\boldsymbol{\Psi}_R^{-1} \\
&= (\mathbf{X} - \mathbf{X}\mathbf{S}^{-1}\,\widehat{\mathbf{U}}^2)\,\mathbf{Q}_{4,R}\,\boldsymbol{\Psi}_R^{-1} \\
&= (\mathbf{X} - \widehat{\mathbf{E}}^{(d)}\widehat{\mathbf{U}})\,\mathbf{Q}_{4,R}\,\boldsymbol{\Psi}_R^{-1}, \qquad (28)
\end{aligned}$$

where we used (22), $\mathbf{X}^T\widehat{\mathbf{E}} = \mathbf{X}^T\widehat{\mathbf{E}}^{(d)}$, and (24). From (28), it can be seen that $\widehat{\mathbf{F}}^{(d)}$ depends on $\mathbf{X}$ and $\widehat{\mathbf{U}}$ only. Hence, like $\widehat{\mathbf{E}}^{(d)}$, the determinate part $\widehat{\mathbf{F}}^{(d)}$ is fixed when $\mathbf{U}$ has been estimated. Note the similarity between the expressions for $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{F}}^{(d)}$ in (22) and (28), respectively. From these, we obtain

$$\begin{aligned}
\widehat{\mathbf{F}}^{(u)} &= \widehat{\mathbf{F}} - \widehat{\mathbf{F}}^{(d)} \\
&= [(\mathbf{X} - \widehat{\mathbf{E}}\widehat{\mathbf{U}}) - (\mathbf{X} - \widehat{\mathbf{E}}^{(d)}\widehat{\mathbf{U}})]\,\mathbf{Q}_{4,R}\,\boldsymbol{\Psi}_R^{-1} \\
&= (\widehat{\mathbf{E}}^{(d)}\widehat{\mathbf{U}} - \widehat{\mathbf{E}}\widehat{\mathbf{U}})\,\mathbf{Q}_{4,R}\,\boldsymbol{\Psi}_R^{-1} \\
&= -\widehat{\mathbf{E}}^{(u)}\widehat{\mathbf{U}}\,\mathbf{Q}_{4,R}\,\boldsymbol{\Psi}_R^{-1}. \qquad (29)
\end{aligned}$$

Hence, the arbitrariness of $\widehat{\mathbf{E}}^{(u)}$ in (26) is directly related to the arbitrariness of $\widehat{\mathbf{F}}^{(u)}$ in (29).

Through a sampling scheme, the expression (29) can be used to obtain a large number of realizations of $\widehat{\mathbf{F}}^{(u)}$ and, hence, of the factor scores $\widehat{\mathbf{F}}$. From these realizations we can estimate probability densities for each factor score in $\widehat{\mathbf{F}}$. These are analogous to posterior densities of $\mathbf{F}$ in the empirical Bayes framework (Skrondal and Rabe-Hesketh, 2004), and have the same mean $\widehat{\mathbf{F}}^{(d)}$ and variance $\text{Var}(\widehat{\mathbf{F}}^{(u)})$ (given by Lemma 3.1(iv)). Note, however, that we do not assume a distributional form while a prior distribution needs to be specified to obtain the posterior densities in the empirical Bayes framework (Bartholomew, 1981).

Under model II, the sampling procedure of $\widehat{\mathbf{F}}^{(u)}$ is as follows. We assume that the algorithm of model II is run and satisfactory estimates of $\mathbf{P}\mathbf{T}^{-T}$ and $\mathbf{U}$ have been obtained, where $\mathbf{T}$ is an orthogonal or oblique rotation matrix. An unrotated realization of $\widehat{\mathbf{F}}$ can be obtained by computing $\widehat{\mathbf{F}}^{(d)}$ as in (28), and computing $\widehat{\mathbf{F}}^{(u)}$ as in (29) by using (26) with a random $\mathbf{M} \in \mathbb{R}^{(N-J-1)\times\tilde{R}}$ such that $\mathbf{M}^T\mathbf{M} = \mathbf{I}_{\tilde{R}}$. To obtain a rotated realization of $\widehat{\mathbf{F}}\mathbf{T}$, we rotate $\widehat{\mathbf{F}} = \widehat{\mathbf{F}}^{(d)} + \widehat{\mathbf{F}}^{(u)}$ by $\mathbf{T}$. After a large number of realizations of $\widehat{\mathbf{F}}\mathbf{T}$ have been obtained in this way, the probability density estimates can be computed for each factor score in $\widehat{\mathbf{F}}\mathbf{T}$ using a density estimation method. To ensure a uniform distribution over all columnwise orthonormal $\mathbf{M}$, it can be taken as the columnwise orthonormal $\mathbf{Q}$ in a QR decomposition of a matrix with random and independent entries sampled from the standard normal distribution, where the corresponding upper triangular $\mathbf{R}$ has positive diagonal entries (Stewart, 1980). The above procedure to obtain factor score densities is used in the application in Section 6.

5. Simulation study

Here, we assess the performance of the algorithms for models I and II in a Monte Carlo simulation study. We compare them to MINRES of Harman and Jones (1966), which fits the correlation model (2). Synthetic data is generated via sampling from a normal distribution using the true correlation matrix (Section 5.1) and by adding a random noise term to the true data factor model (Section 5.2). We focus on retrieval of true loadings $\mathbf{P}$, of the true unique standard deviations $\mathbf{U}$, and of the true determinate factor part $\mathbf{F}^{(d)}$. The ALS algorithm for model I is run 10 times with random starting values for $\mathbf{F}$ and $\mathbf{E}$ (satisfying the model assumptions), and convergence criterion 10^{-6}. The solution with the smallest $\text{ssq}(\mathbf{X} - \mathbf{F}\,\mathbf{P}^T - \mathbf{E}\,\mathbf{U})$ is kept.

Note that when generating data by using the true correlation matrix (2), the factor scores are necessarily left out of consideration. The purpose of the simulation study in Section 5.1 is to compare the estimation accuracy of only the loadings and unique variances when the data is generated as usual in factor analysis simulation studies. In Section 5.2 the data is generated as the true data factor model (4) plus a noise term, which allows to also assess the estimation accuracy of $\mathbf{F}^{(d)}$. Note, however, that the error structure in both simulation methods is different.

Estimation accuracy is evaluated by means of the mean absolute error (MAE) and root mean square error (RMSE), which are defined for estimate $\widehat{\mathbf{P}}$ and true value $\mathbf{P}$ as

$$\text{MAE}(\mathbf{P}) = (JR)^{-1}\sum_{j=1}^{J}\sum_{r=1}^{R}|p_{jr} - \hat{p}_{jr}|, \qquad \text{RMSE}(\mathbf{P}) = \sqrt{(JR)^{-1}\sum_{j=1}^{J}\sum_{r=1}^{R}(p_{jr} - \hat{p}_{jr})^2}. \qquad (30)$$

The MAE and RMSE for $\mathbf{U}$ and $\mathbf{F}^{(d)}$ are defined analogously.

For the implementation of MINRES, Harman and Jones (1966) proposed three algorithms. We use the algorithm in which $\mathbf{P}$ is updated for each row separately. This makes it possible to prevent so-called Heywood cases in which an estimated communality is larger than 1, implying a negative unique variance. In MINRES the unique variances are set equal to $\text{diag}(\mathbf{S})-\text{diag}(\mathbf{P}\mathbf{P}^T)$, after $\mathbf{P}$ has been estimated. For updating $\mathbf{P}$, we fit $\mathbf{P}\mathbf{P}^T$ to $\mathbf{S}-\mathbf{I}_J$. For row $\mathbf{p}_j^T$ of $\mathbf{P}$, the update is as follows. Let $\mathbf{s}_j$ be the jth column of $\mathbf{S}-\mathbf{I}_J$, and let $\mathbf{P}_j$ be matrix $\mathbf{P}$ with row j replaced by zeros. Vector $\mathbf{p}_j$ is found by minimizing $\text{ssq}(\mathbf{s}_j-\mathbf{P}_j\,\mathbf{p}_j)$. This yields $\mathbf{p}_j=(\mathbf{P}_j^T\mathbf{P}_j)^{-1}\mathbf{P}_j^T\mathbf{s}_j$. When $\mathbf{p}_j^T\mathbf{p}_j>1$ (a Heywood case), the regression is performed under the constraint $\mathbf{p}_j^T\mathbf{p}_j=1$ by using the method of Browne (1967); see also Ten Berge and Nevels (1977). The MINRES algorithm is run 10 times with random starting values for $\mathbf{P}$ and convergence criterion 10^{-6}. The solution with the smallest $\text{ssq}(\mathbf{S}-\mathbf{P}\mathbf{P}^T-\mathbf{U}^2)$ is kept.

Note that Heywood cases do not occur in model II, since MRFA is used to estimate $\mathbf{U}$ and the explained common variances (diagonal of $\widehat{\mathbf{P}}\widehat{\mathbf{P}}^T$) are guaranteed to not exceed the communalities (diagonal of $\mathbf{S}-\widehat{\mathbf{U}}^2$). Under model I, the update of $\mathbf{U}$ is not directly related to the update of $\mathbf{P}$ as in MINRES. Although it may happen that the diagonal of $\widehat{\mathbf{P}}\widehat{\mathbf{P}}^T$ contains entries larger than 1, this does not present a problem in the estimation of $\mathbf{U}$. Such cases are not corrected.

5.1. Generating data via the true correlation model

We make synthetic data with $J=9$ observed variables and $R=3$ orthogonal factors. As true parameter values, we set

$$\mathbf{P}=\begin{bmatrix}0.9 & 0 & 0\\ 0 & 0.8 & 0\\ 0 & 0 & 0.5\\ 0.7 & 0.6 & 0\\ 0.7 & 0 & 0.3\\ 0 & 0.6 & 0.3\\ 0.3 & 0.2 & 0.1\\ 0.6 & 0.5 & 0.3\\ 0.6 & 0.6 & 0.4\end{bmatrix},\qquad \text{diagonal of }\mathbf{U}^2=\begin{pmatrix}0.19\\ 0.36\\ 0.75\\ 0.15\\ 0.42\\ 0.55\\ 0.86\\ 0.30\\ 0.12\end{pmatrix}, \tag{31}$$

where the true correlation matrix $\mathbf{S}=\mathbf{P}\mathbf{P}^T+\mathbf{U}^2$ has ones on the diagonal. Hence, the data feature both high and low communalities. The total common variance equals $\text{trace}(\mathbf{S}-\mathbf{U}^2)=5.3$, which is 58.9% of the total variance. The common variance due to each factor (on the diagonal of $\mathbf{P}^T\mathbf{P}$) is 2.60, 2.01, and 0.69, which is 49.1%, 37.9%, and 13.0% of the total common variance, and 28.9%, 22.3%, and 7.7% of the total variance, respectively. The minimal correlation per factor is 0.77, 0.65, and 0.16. Hence, the third factor is considerably weaker than the first two factors. For $\text{Var}(\mathbf{F}^{(d)})=\mathbf{P}^T\mathbf{S}^{-1}\mathbf{P}$ and $\text{Var}(\mathbf{F}^{(u)})=\mathbf{I}_3-\mathbf{P}^T\mathbf{S}^{-1}\mathbf{P}$ we obtain

$$\text{Var}(\mathbf{F}^{(d)})=\begin{bmatrix}0.88 & 0.07 & 0.05\\ 0.07 & 0.82 & 0.09\\ 0.05 & 0.09 & 0.58\end{bmatrix},\qquad \text{Var}(\mathbf{F}^{(u)})=\begin{bmatrix}0.12 & -0.07 & -0.05\\ -0.07 & 0.18 & -0.09\\ -0.05 & -0.09 & 0.42\end{bmatrix}. \tag{32}$$

Logically, a larger uncertainty due to indeterminacy (larger $\text{Var}(\mathbf{f}_r^{(u)})=1-\text{Var}(\mathbf{f}_r^{(d)})$) implies a smaller minimal correlation (given by $2\,\text{Var}(\mathbf{f}_r^{(d)})-1$).

The data is generated as $\mathbf{Z}=\mathbf{Y}\mathbf{S}^{1/2}$, where $\mathbf{Y}$ is $N\times J$ with each entry randomly sampled from the standard normal distribution. We use both $N=100$ and $N=500$ as sample size. Models I and II are applied to standardized $\mathbf{Z}$. MINRES is applied to the correlation matrix of $\mathbf{Z}$. We use $R=3$ factors in the estimation procedures. Since we do not have true values for $\mathbf{F}^{(d)}$, we focus on retrieval of the true $\mathbf{P}$ and $\mathbf{U}$ only. For each sample size N, we generate 500 datasets $\mathbf{Z}$ and apply the three algorithms. After convergence, orthogonal Procrustes rotation (Schönemann, 1966) is applied to the estimates of $\mathbf{P}$ to approximate the true $\mathbf{P}$ as close as possible. In Table 1 we report the mean and standard deviation of the MAE and RMSE for $\mathbf{P}$ and $\mathbf{U}$ for each sample size N. The performance of the three methods is very similar, although model II has slightly worse performance in all cells. This is probably due to the additional constraint $N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{E}\,\mathbf{U})=\mathbf{O}$ in model II that is not present in model I and MINRES. Although this constraint holds in the population, it may not hold in the sample.

5.2. Generating data via the true data factor model

Here, we perform Monte Carlo simulations in which the data is generated as $\mathbf{Z}=\mathbf{F}\mathbf{P}^T+\mathbf{E}\,\mathbf{U}+\sigma\,\mathbf{Y}$, with $\mathbf{P}$ and $\mathbf{U}$ as in (31), random $\mathbf{F}$ and $\mathbf{E}$ satisfying the model assumptions $N^{-1}[\mathbf{F}\ \mathbf{E}]^T[\mathbf{F}\ \mathbf{E}]=\mathbf{I}_{R+J}$, $\mathbf{F}^T\mathbf{1}_N=\mathbf{0}$, and $\mathbf{E}^T\mathbf{1}_N=\mathbf{0}$, and $\mathbf{Y}$ with entries randomly sampled from the standard normal distribution. The parameter $\sigma>0$ indicates the strength of the noise term, and we use $\sigma=0.4$ and $\sigma=0.8$ for low and high noise situations, respectively. For the same $\mathbf{F}$, $\mathbf{E}$, and σ, we generate 500 datasets $\mathbf{Z}$. As above, we consider sample sizes $N=100$ and $N=500$. Models I and II are applied to standardized $\mathbf{Z}$, and MINRES is applied to the correlation matrix $\mathbf{S}_Z$ of $\mathbf{Z}$. We consider retrieval of the true $\mathbf{P}$ and $\mathbf{U}$, and of the determinate factor part $\mathbf{F}^{(d)}$. For MINRES, we take $\mathbf{Z}\mathbf{S}_Z^{-1}\widehat{\mathbf{P}}$ as estimate of $\mathbf{F}^{(d)}$. This is the least squares estimate when $\widehat{\mathbf{P}}=N^{-1}\mathbf{Z}^T\widehat{\mathbf{F}}$. For all three methods, orthogonal Procrustes rotation is applied to the estimates $\widehat{\mathbf{P}}$ to approximate the true $\mathbf{P}$ as close as possible. The

Table 1
Means and standard deviations of MAE and RMSE values for estimates of **P** and **U** when generating 500 datasets via the true correlation model, with sample size N.

N	Method	**P** estimates		**U** estimates	
		MAE	RMSE	MAE	RMSE
100	Model I	0.06 (0.01)	0.08 (0.02)	0.07 (0.02)	0.09 (0.05)
100	Model II	0.07 (0.02)	0.09 (0.03)	0.12 (0.04)	0.16 (0.06)
100	MINRES	0.06 (0.01)	0.08 (0.02)	0.07 (0.03)	0.10 (0.06)
500	Model I	0.03 (0.01)	0.03 (0.01)	0.03 (0.01)	0.03 (0.01)
500	Model II	0.03 (0.01)	0.04 (0.01)	0.05 (0.01)	0.06 (0.02)
500	MINRES	0.03 (0.01)	0.04 (0.01)	0.03 (0.01)	0.03 (0.01)

Table 2
Means and standard deviations of MAE and RMSE values for estimates of **P**, **U**, and $\mathbf{F}^{(d)}$ when generating 500 datasets via the true data factor model, with sample size N and noise variance σ^2.

(N, σ)	Method	**P** estimates		**U** estimates		$\mathbf{F}^{(d)}$ estimates	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
(100, 0.4)	Model I	0.05 (0.01)	0.07 (0.02)	0.08 (0.02)	0.10 (0.04)	0.27 (0.04)	0.35 (0.06)
(100, 0.4)	Model II	0.06 (0.01)	0.07 (0.02)	0.08 (0.03)	0.11 (0.07)	0.29 (0.04)	0.37 (0.07)
(100, 0.4)	MINRES	0.05 (0.01)	0.07 (0.02)	0.08 (0.02)	0.11 (0.04)	0.27 (0.03)	0.35 (0.05)
(100, 0.8)	Model I	0.13 (0.02)	0.16 (0.03)	0.20 (0.05)	0.25 (0.06)	0.49 (0.06)	0.63 (0.08)
(100, 0.8)	Model II	0.13 (0.02)	0.16 (0.03)	0.17 (0.05)	0.23 (0.07)	0.49 (0.06)	0.64 (0.08)
(100, 0.8)	MINRES	0.13 (0.02)	0.16 (0.03)	0.21 (0.05)	0.25 (0.07)	0.48 (0.06)	0.62 (0.09)
(500, 0.4)	Model I	0.03 (0.00)	0.04 (0.00)	0.07 (0.01)	0.09 (0.01)	0.24 (0.01)	0.30 (0.01)
(500, 0.4)	Model II	0.03 (0.00)	0.04 (0.01)	0.06 (0.01)	0.07 (0.01)	0.24 (0.01)	0.30 (0.01)
(500, 0.4)	MINRES	0.03 (0.00)	0.04 (0.00)	0.07 (0.01)	0.09 (0.01)	0.24 (0.01)	0.30 (0.01)
(500, 0.8)	Model I	0.09 (0.01)	0.11 (0.01)	0.18 (0.02)	0.21 (0.02)	0.40 (0.03)	0.50 (0.04)
(500, 0.8)	Model II	0.09 (0.01)	0.11 (0.02)	0.16 (0.03)	0.19 (0.04)	0.41 (0.03)	0.52 (0.05)
(500, 0.8)	MINRES	0.09 (0.01)	0.11 (0.01)	0.18 (0.02)	0.21 (0.03)	0.40 (0.02)	0.50 (0.03)

rotated $\widehat{\mathbf{P}}\mathbf{T}$ and the rotated $\widehat{\mathbf{F}}^{(d)}\mathbf{T}$ are then evaluated in terms of MAE and RMSE. In Table 2 we report the mean and standard deviation of the MAE and RMSE for **P**, **U**, and $\mathbf{F}^{(d)}$ for each sample size N and each noise level σ.

Compared to Table 1, the performance of the methods is still very similar. Here, model II does not always have the worst performance. On the contrary, in the high noise case ($\sigma = 0.8$) the estimation of **U** by model II is better than that of the other two methods. The MAE and RMSE values for $\mathbf{F}^{(d)}$ may seem rather large compared to those for **P** and **U**. However, $\mathbf{F}^{(d)}$ has values on a larger scale than **P** and **U**. Moreover, the Procrustes rotation matches the estimate $\widehat{\mathbf{P}}$ to the true **P**, which may not be optimal for estimate $\widehat{\mathbf{F}}^{(d)}$ and its true value $\mathbf{F}^{(d)}$.

We conclude that, despite model II having an additional constraint, its ability to recover underlying loadings, unique variances, and determinate factor parts is similar to that of model I. Hence, the advantage of being able to compute the explained common variance is not counterweighted by much worse recovery performance. Both data factor models perform similar to MINRES, which does not offer a description of the complete class of solutions for $\widehat{\mathbf{F}}$ (and $\widehat{\mathbf{E}}$), nor the ability to compute the explained common variance.

6. Application: WISC-III

Here, we apply model II to a dataset of scores of $N = 280$ children on the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991), where the children were diagnosed with learning disabilities. The dataset is used in a cross-validation study by Grice et al. (1999) and part of the dataset (215 children) is used in Grice (2001) to illustrate various methods of factor score estimation. The dataset was obtained from the website of James W. Grice, Oklahoma State University.[1]

The WISC-III consists of 13 subtests, $J = 12$ of which are contained in the dataset. Observed scores range from 1 to 19 for each subtest and model II applied to the standardized dataset. As in Grice (2001) and consistent with the WISC-III manual, we extract $R = 4$ factors and apply Varimax rotation to the obtained factors $\widehat{\mathbf{F}}$ and loadings $\widehat{\mathbf{P}}$. The rotated loadings matrix is given in Table 3. The rotated loadings are very similar to the solution obtained in Grice (2001), with the same loadings being larger than 0.4. Finite sample standard errors for the estimated rotated loadings and unique variances are obtained via a bootstrap procedure with 2000 bootstrap samples (e.g., Zhang, 2014). The bootstrap standard errors are also given in Table 3. The size of the standard errors is similar to those reported in Zhang (2014), with factor 4 having the largest values.

[1] http://psychology.okstate.edu/faculty/jgrice/factorscores/RawData.sas.

Table 3
Varimax rotated loadings, unique variances, and explained common variances when applying model II to the 12 subtests of the WISC-III dataset. Bootstrap standard errors are given in round brackets.

Subtest	Factor loadings				Unique Var	ECV%
	1	2	3	4		
Picture completion	0.26 (0.05)	**0.62** (0.06)	0.16 (0.06)	0.03 (0.08)	0.48 (0.08)	91.7
Information	**0.75** (0.06)	0.13 (0.05)	0.09 (0.06)	0.14 (0.09)	0.32 (0.08)	90.1
Coding	−0.02 (0.05)	0.15 (0.06)	**0.67** (0.10)	0.08 (0.08)	0.48 (0.13)	92.0
Similarities	**0.71** (0.05)	0.20 (0.05)	0.01 (0.05)	−0.00 (0.08)	0.42 (0.08)	93.9
Picture arrangement	0.27 (0.06)	**0.43** (0.08)	**0.44** (0.09)	0.01 (0.14)	0.47 (0.08)	84.6
Arithmetic	0.32 (0.09)	0.04 (0.07)	0.30 (0.10)	**0.52** (0.21)	0.48 (0.16)	88.2
Block design	0.22 (0.05)	**0.72** (0.07)	0.24 (0.06)	0.22 (0.07)	0.30 (0.08)	97.1
Vocabulary	**0.79** (0.04)	0.21 (0.04)	0.10 (0.05)	0.07 (0.06)	0.32 (0.07)	99.2
Object assembly	0.13 (0.04)	**0.76** (0.07)	0.08 (0.05)	−0.09 (0.08)	0.35 (0.11)	93.9
Comprehension	**0.61** (0.05)	0.14 (0.06)	0.14 (0.06)	0.24 (0.09)	0.50 (0.07)	93.7
Symbol search	0.14 (0.05)	0.19 (0.05)	**0.77** (0.11)	0.04 (0.09)	0.31 (0.13)	93.5
Digit span	0.16 (0.07)	0.08 (0.06)	−0.04 (0.07)	**0.55** (0.22)	0.62 (0.23)	87.8

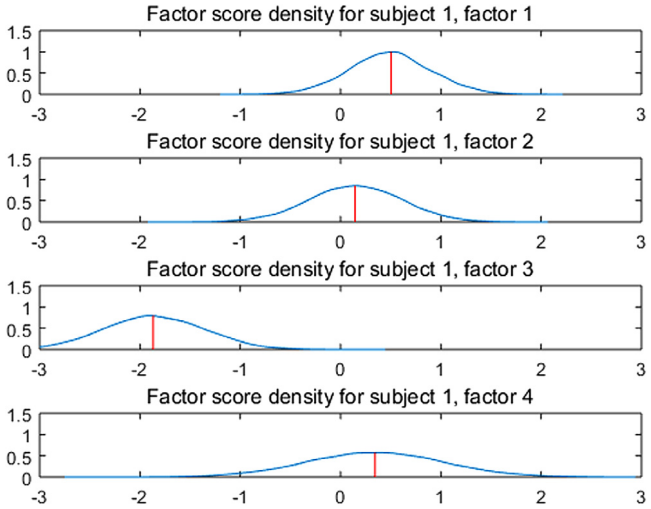

Fig. 1. Factor score density estimates for the first row of $\widehat{\mathbf{F}}$, based on 10 000 realizations, when fitting model II to the WISC-III dataset. The red lines indicate the values of $(\widehat{\mathbf{F}}^{(d)})_{1,r}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For $\mathrm{Var}(\widehat{\mathbf{F}}^{(d)})$ and $\mathrm{Var}(\widehat{\mathbf{F}}^{(u)})$ we obtain

$$\mathrm{Var}(\widehat{\mathbf{F}}^{(d)}) = \begin{bmatrix} 0.82 & 0.07 & 0.02 & 0.08 \\ 0.07 & 0.78 & 0.08 & -0.00 \\ 0.02 & 0.08 & 0.74 & 0.05 \\ 0.08 & -0.00 & 0.05 & 0.53 \end{bmatrix}, \quad \mathrm{Var}(\widehat{\mathbf{F}}^{(u)}) = \begin{bmatrix} 0.18 & -0.07 & -0.02 & -0.08 \\ -0.07 & 0.22 & -0.08 & 0.00 \\ -0.02 & -0.08 & 0.26 & -0.05 \\ -0.08 & 0.00 & -0.05 & 0.47 \end{bmatrix}. \tag{33}$$

The total explained common variance is 92.5%, and the contribution of each factor is 34.8%, 26.7%, 20.8%, and 10.2%, respectively. The minimal correlation for each factor is 0.64, 0.55, 0.48, and 0.05. As explained in Section 3, the minimal correlation is smaller for factors with smaller explained common variance. Factor 4 has minimal correlation close to zero. Hence, interpretation of this factor is highly dubious. Since the explained common variance is very high for each subtest (see Table 3), we also try extracting $R = 3$ factors. In that case, the total explained common variance is 86.0% and the minimal correlations are 0.68, 0.56, and 0.48, respectively, which is acceptable. Hence, a solution with $R = 3$ factors may be preferred for the WISC-III dataset instead.

Using the procedure outlined in Section 4.4, we generate 10 000 realizations of the Varimax rotated factor scores matrix for the $R = 4$ solution. For subject 1, we estimate the factor score densities due to factor indeterminacy by applying Gaussian kernel density estimation to the first row of the sampled factor scores matrices. The densities are depicted in Fig. 1. As can be seen, the density of factor 4 is very flat and yields little information on the factor score.

Next, we consider the distribution of the sampled factor scores for all subjects together. For each factor we have 10 000 times 280 sampled scores. In Fig. 2 the boxplots of the scores are depicted for each factor, with the boxes indicating the central 90% of the scores. The four distributions have skewness 0.18, −0.06, 0.13, and 0.04, respectively, and kurtosis 2.83, 2.98, 3.11, and 2.93, respectively. Although these values suggest a shape close to the normal distribution, QQ-plots of the sampled factor scores show that the tails of the empirical distributions deviate significantly from the normal tails for some factors (results not depicted). Note that normality of the sampled factor score distribution of all subjects together does

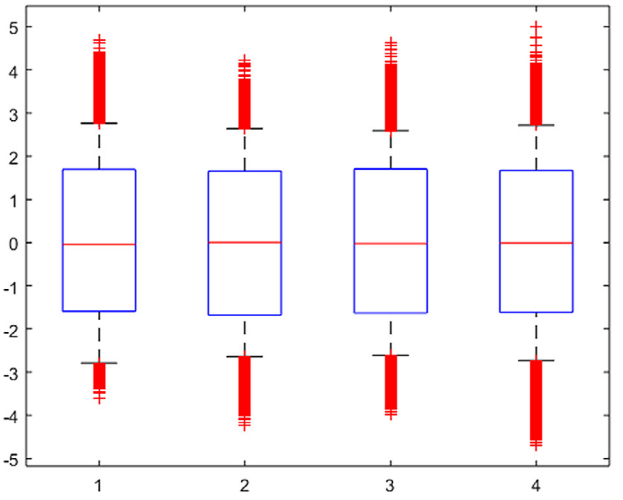

Fig. 2. Boxplots of 10 000 realizations of $\widehat{\mathbf{F}}$ for all subjects together, for each factor separately, when fitting model II to the WISC-III dataset. The box limits are the 5th and 95th percentiles, respectively.

not follow from normality of the sampled factor score distribution for each subject separately (since these have different means for different subjects). Normality of the latter distributions holds in the empirical Bayes framework when assuming a normal prior distribution. For the sampled factor scores under the data factor models, however, normality of the sampled factor scores for each subject separately does not need to hold.

In the above, we fitted data factor model II to the WISC-III dataset. When model I is fitted instead, the estimated loadings are very similar but the estimated unique variances are somewhat larger. The bootstrap standard errors are also very similar. The matrix $\mathbf{S} - \widehat{\mathbf{U}}^2$ has both positive and negative eigenvalues and, hence, it is not possible to compute a percentage of explained common variance.

7. Discussion

In this paper we have considered data factor models, in which the factor loadings $\mathbf{P}$, unique variances $\mathbf{U}^2$, factor scores $\mathbf{F}$, and unique parts $\mathbf{E}$ are all parameters to be estimated. Our new data factor model II is equal to the existing data factor model I (Sočan, 2003; De Leeuw, 2004) with the additional constraint $N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{E}\,\mathbf{U}) = \mathbf{O}$ to obtain a rigorous distinction between the common and unique parts. The parameters of models I and II are estimated simultaneously. Estimation of model I is done iteratively via alternating least squares (Sočan, 2003; De Leeuw, 2004). The estimation procedure for model II is noniterative and combines estimation of the loadings and unique variances by MRFA of Ten Berge and Kiers (1991) with unweighted least squares estimation of the factor scores and unique parts. By using MRFA the explained common variance can be computed, which is important when choosing the number of factors. This is not possible under model I or other commonly used factor methods such as MINRES or MLFA. Although the estimated factors under models I and II are orthogonal, after estimation they can be rotated orthogonally or obliquely together with the estimated loadings without loss of fit. The simulation study shows that the additional constraint $N^{-1}\mathbf{E}^T(\mathbf{X} - \mathbf{E}\,\mathbf{U}) = \mathbf{O}$ in model II does not result in much worse estimation accuracy compared to model I and MINRES. Additionally, the simulation study in Sočan (2003) shows that the accuracy of MRFA loading estimates is similar to or better than MINRES and MLFA. Matlab codes for estimating the parameters of data factor models I and II, together with a short user guide and the dataset from Grice (2001), are available via the author's website.[2]

For models I and II, we obtain a decomposition of the estimated factor scores $\widehat{\mathbf{F}}$ and unique parts $\widehat{\mathbf{E}}$ into determinate and indeterminate parts by extending the classical analysis of Guttman (1955) to the case of imperfect fit. Factor indeterminacy can be reported via minimal correlations, factor score standard deviations due to indeterminacy, and factor score density estimates obtained by uniform sampling of the indeterminate part. These densities can be used in practice as a visual tool to assess the uncertainty in the factor scores due to indeterminacy. This is analogous to the empirical Bayes framework of factor score estimation, but without the need of specifying a prior distribution for the factor scores.

Contrary to factor score estimates of the form $\widehat{\mathbf{F}} = \mathbf{X}\mathbf{B}$, the estimates $(\widehat{\mathbf{F}}, \widehat{\mathbf{E}})$ under models I and II satisfy the usual assumptions of the common factor model: $N^{-1}\widehat{\mathbf{F}}^T\widehat{\mathbf{E}} = \mathbf{O}$, $N^{-1}\widehat{\mathbf{E}}^T\widehat{\mathbf{E}}$ diagonal, $N^{-1}\mathbf{X}^T\widehat{\mathbf{F}} = \widehat{\mathbf{P}}\,\widehat{\mathbf{\Phi}}$, and $N^{-1}\widehat{\mathbf{F}}^T\widehat{\mathbf{F}} = \mathbf{\Phi}$. Moreover, the estimates $(\widehat{\mathbf{F}}, \widehat{\mathbf{E}})$ represent the complete class of solutions $(\widehat{\mathbf{F}}, \widehat{\mathbf{E}})$ satisfying these assumptions.

We conclude with some directions for future research. Although finite sample standard errors for the estimated loadings, unique variances, and factor correlations can be obtained via a bootstrap procedure (as demonstrated in the application),

[2] http://www.alwinstegeman.nl/matlab.htm.

asymptotic standard errors under normality are currently only available for the unique variance estimates by MRFA under model II (Shapiro and Ten Berge, 2002). Asymptotic standard errors could also be derived for the estimates under model I, and possibly also for the estimated loadings under model II. The latter may be a difficult problem though.

The data factor models do not assume normally distributed observed variables or factors. As such they may be especially suitable for nonnormal data. Other approaches exist for nonnormal data such as the skew-normal factor model (Montanari and Viroli, 2010), the skew-t factor model (e.g., Lin et al., 2015) or nonlinear factor models. A drawback of the skew-normal factor model is that the range of the skewness of the modelled common part is rather limited. In Smits et al. (in press) it is shown that for each item the common part under the skew-normal factor model can be approximated up to its third moments by that of a quadratic factor model, while the converse is not true due to the skewness limitation of the skew-normal factor model. The quadratic factor model is more difficult to interpret than a linear factor model, however. The data factor models may form an appropriate alternative when their performance on synthetic and real life nonnormal data is good compared to factor models assuming a skewed factor distribution.

Another subject for further research is how to handle missing data in the data factor models. In probabilistic factor models missing data (when missing at random) can be handled via multiple imputation in a maximum likelihood estimation framework (e.g., Schafer and Graham, 2002). The data factor models are not probabilistic, however, but are similar to component models such as principal component analysis (PCA). Handling of missing data may be possible by adapting an algorithm for missing data in PCA, such as those presented in Grung and Manne (1998).

Finally, the possibility of extending model II to the case of more observed variables than observations may be considered. For model I this has been done by Trendafilov and Unkel (2011). Some difficulties need to be overcome when adopting their approach for model II. First, the data correlation matrix $\mathbf{S}$ is singular and, hence, its inverse cannot be used in the algorithm. Second, the constraints $N^{-1}\mathbf{E}^T\mathbf{E} = \mathbf{I}_J$ and $N^{-1}\mathbf{E}^T(\mathbf{X} - \mathbf{E}\,\mathbf{U}) = \mathbf{O}$ cannot be fulfilled. Trendafilov and Unkel (2011) replace $N^{-1}\mathbf{E}^T\mathbf{E} = \mathbf{I}_J$ with the more general constraint $N^{-1}\mathbf{E}^T\mathbf{E}\,\mathbf{U} = \mathbf{U}$, which implies that the correlation model $\mathbf{S} = \mathbf{P}\,\mathbf{\Phi}\,\mathbf{P}^T + \mathbf{U}^2$ still holds. For the constraint $N^{-1}\mathbf{E}^T(\mathbf{X} - \mathbf{E}\,\mathbf{U}) = \mathbf{O}$ a similar alternative should be found. Moreover, the rigorous distinction between the common and unique parts should be kept, such that the explained common variance can still be computed.

Acknowledgements

The author would like to thank Henk Kiers (University of Groningen) for commenting on earlier versions of this manuscript, Jeroen Vermunt (Tilburg University) for pointing out the link with the empirical Bayes approach of factor score estimation, and the associate editor and anonymous reviewers for bringing to the author's attention several related publications on factor analysis methods, and providing valuable suggestions that have improved the accessibility of the manuscript.

Appendix. Remarks on the data factor models

Here we gather some technical remarks on the data factor models I and II.

1. The factor score estimates $\widehat{\mathbf{F}}$ under models I and II (rotated or not) and under the maximum likelihood ratio estimation of the fixed factor model (McDonald, 1979) are partly indeterminate but satisfy the properties of factor scores under the common factor model: $N^{-1}\widehat{\mathbf{F}}^T\widehat{\mathbf{F}} = \widehat{\mathbf{\Phi}}$, $N^{-1}\widehat{\mathbf{F}}^T\widehat{\mathbf{E}} = \mathbf{O}$, and $N^{-1}\mathbf{X}^T\widehat{\mathbf{F}} = \widehat{\mathbf{P}}\,\widehat{\mathbf{\Phi}}$. On the other hand, estimates of the form $\widehat{\mathbf{F}} = \mathbf{X}\,\mathbf{B}$, although fully determinate, do not satisfy all these assumptions (Schönemann and Wang, 1972). Models I and II are the only factor models offering an expression for the complete class of solutions for $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{E}}$.
2. Alternative estimates $(\widehat{\mathbf{F}}, \widehat{\mathbf{E}})$ and $(\widetilde{\mathbf{F}}, \widetilde{\mathbf{E}})$ with the same determinate parts but different indeterminate parts have different properties under models I and II. For fixed $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{U}}$, it holds that $\mathrm{ssq}(\mathbf{X} - \widehat{\mathbf{F}}\widehat{\mathbf{P}}^T - \widehat{\mathbf{E}}\widehat{\mathbf{U}}) = \mathrm{ssq}(\mathbf{X} - \widetilde{\mathbf{F}}\widehat{\mathbf{P}}^T - \widetilde{\mathbf{E}}\widehat{\mathbf{U}})$ under both models. However, under model I the fitted model parts are also the same: $\widehat{\mathbf{F}}\widehat{\mathbf{P}}^T + \widehat{\mathbf{E}}\widehat{\mathbf{U}} = \widetilde{\mathbf{F}}\widehat{\mathbf{P}}^T + \widetilde{\mathbf{E}}\widehat{\mathbf{U}}$. This follows directly from the ALS updates in Section 4.2. Under model II we generally have $\widehat{\mathbf{F}}\widehat{\mathbf{P}}^T + \widehat{\mathbf{E}}\widehat{\mathbf{U}} \neq \widetilde{\mathbf{F}}\widehat{\mathbf{P}}^T + \widetilde{\mathbf{E}}\widehat{\mathbf{U}}$, due to the additional constraint $N^{-1}\mathbf{E}^T(\mathbf{X} - \mathbf{E}\,\mathbf{U}) = \mathbf{O}$.
3. A model with a similar aim as model II has been proposed by Kiers in Sočan (2003) and equals model I with the additional constraint of $N^{-1}\mathbf{E}^T\mathbf{X}$ being diagonal. Together with $N^{-1}\mathbf{E}^T\mathbf{E} = \mathbf{I}_J$, this implies that the unique part of variable j is uncorrelated to the common parts of variables $k \neq j$ (i.e., columns $k \neq j$ of $\mathbf{X} - \mathbf{EU}$). However, the unique part of variable j may still be correlated to the common part of variable j. This can only be solved by requiring that $N^{-1}\mathbf{E}^T\mathbf{X} = \mathbf{U}$, which is equivalent to $N^{-1}\mathbf{E}^T(\mathbf{X} - \mathbf{EU}) = \mathbf{O}$ as is the constraint under model II. Below, we refer to this second model in Sočan (2003) as the intermediate model.
4. Model I and the intermediate model in Sočan (2003) are used in a simulation study of Ng (2015) to assess the accuracy of factor score estimation. However, factor indeterminacy is not explicitly mentioned nor explicitly taken into account when evaluating the estimation accuracy in Ng (2015). As criterion the R^2 of the regression of each estimated factor on a constant and all true factors is used. It would be better to explicitly separate the determinate and indeterminate factor parts, since only the determinate parts can be accurately estimated. In our simulation study in Section 5 we are able to do this, based on the analysis of the estimate of $\mathbf{F}$ under models I and II. Ng (2015) reports that the ALS algorithm to fit the intermediate model is computationally demanding and sometimes fails to converge. In our simulation study the algorithm for model II converges fast in all cases.

5. As mentioned in Section 4.4, the factor score estimates under models I and II have a link with the empirical Bayes framework of factor score estimation (Skrondal and Rabe-Hesketh, 2004). In the latter, $\widehat{\mathbf{F}}^{(d)}$ is the mean and $\mathrm{Var}(\widehat{\mathbf{F}}^{(u)})$ the covariance matrix of the posterior factor score distribution, given the data $\mathbf{X}$ and estimates $\widehat{\mathbf{P}}$, $\widehat{\mathbf{U}}$, and $\widehat{\mathbf{\Phi}}$. The sampling uncertainty of the factor score estimate $\widehat{\mathbf{F}}^{(d)}$ is given by $\mathrm{Var}(\widehat{\mathbf{F}}^{(d)})$. In the empirical Bayes framework the estimates $\widehat{\mathbf{P}}$, $\widehat{\mathbf{U}}$, and $\widehat{\mathbf{\Phi}}$ are treated as fixed, and the posterior density is not truly posterior in the Bayesian sense. An expression of the sampling uncertainty $\mathrm{Var}(\widehat{\mathbf{F}})$ for an estimate $\widehat{\mathbf{F}} = \mathbf{X}\mathbf{B}$ that includes the estimation uncertainty of the loadings, unique variances, and factor correlations can be found in Hoshino and Shigemasu (2008).

References

Bartholomew, D.J., 1981. Posterior analysis of the factor model. Psychometrika 34, 93–99.
Browne, M.W., 1967. On oblique Procrustes rotation. Psychometrika 32, 125–132.
Browne, M.W., 2001. An overview of analytic rotation in exploratory factor analysis. Multivariate Behav. Res. 36, 111–150.
De Leeuw, J., 2004. Least squares optimal scaling of partially observed linear systems. In: van Montfort, K., Oud, J., Satorra, A. (Eds.), Recent Developments on Structural Equation Models: Theory and Applications. Kluwer Academic Publishers, Dordrecht, pp. 121–134.
De Leeuw, J., 2008. Factor analysis as matrix decomposition. Preprint series: Department of Statistics, University of California, Los Angeles. Available online at: http://www.stat.ucla.edu/~deleeuw/janspubs/2008/notes/deleeuw_U_08g.pdf.
Eckart, C., Young, G., 1936. The approximation of one matrix by another of lower rank. Psychometrika 1, 211–218.
Grice, J.W., 2001. Computing and evaluating factor scores. Psychol. Methods 6, 430–450.
Grice, J.W., Krohn, E.J., Logerquist, S., 1999. Cross-validation of the WISC-III factor structure in two samples of children with learning disabilities. J. Psychoeduc. Assess. 17, 236–248.
Grung, B., Manne, R., 1998. Missing values in principal component analysis. Chemometr. Intell. Lab. Syst. 42, 125–139.
Guttman, L., 1955. The determinacy of factor score matrices with implications for five other basic problems of common factor theory. British J. Statist. Psych. 8, 65–81.
Harman, H.H., Jones, W.H., 1966. Factor analysis by minimizing residuals (MINRES). Psychometrika 31, 351–368.
Hoshino, T., Shigemasu, K., 2008. Standard errors of estimated latent variable scores with estimated structural parameters. Appl. Psychol. Meas. 32, 181–189.
Jöreskog, K.G., 1967. Some contributions to maximum likelihood factor analysis. Psychometrika 32, 443–482.
Lin, T.-I., Wu, P.H., McLachlan, G.J., Lee, S.X., 2015. A robust factor analysis model using the restricted skew-t distribution. TEST 24, 510–531.
McDonald, R.P., 1979. The simultaneous estimation of factor loadings and scores. Br. J. Math. Stat. Psychol. 32, 212–228.
Maraun, M.D., 1996. Metaphor taken as math: indeterminacy in the factor analysis model, with comments and reply. Multivariate Behav. Res. 31, 517–689.
Montanari, A., Viroli, C., 2010. A skew-normal factor model for the analysis of student satisfaction towards university courses. J. Appl. Stat. 37, 473–487.
Mulaik, S.A., 2010. Foundations of Factor Analysis, second ed. Chapman & Hall, FL, USA.
Ng, S., 2015. Constructing common factors from continuous and categorical data. Econometric Rev. 34, 1141–1171.
Schafer, J.L., Graham, J.W., 2002. Missing data: our view of the state of the art. Psychol. Methods 7, 147–177.
Schönemann, P.H., 1966. A generalized solution of the orthogonal Procrustes problem. Psychometrika 31, 1–10.
Schönemann, P.H., Wang, M.-M., 1972. Some new results on factor indeterminacy. Psychometrika 37, 61–91.
Shapiro, A., Ten Berge, J.M.F., 2002. Statistical inference of minimum rank factor analysis. Psychometrika 67, 79–94.
Skrondal, A., Rabe-Hesketh, S., 2004. Generalized Latent Variable Modeling. Chapman & Hall, Boca Raton, USA.
Smits, I.A.M., Timmerman, M.E., Stegeman, A., 2015. Modelling non-normal data: the relationship between the skew-normal factor model and the quadratic factor model. British J. Math. Statist. Psych. http://dx.doi.org/10.1111/bmsp.12062. in press. Available online via.
Sočan, G., 2003. The incremental value of minimum rank factor analysis (PhD Thesis), University of Groningen, The Netherlands.
Spearman, C., 1904. General intelligence objectively determined and measured. Am. J. Psychol. 15, 201–293.
Steiger, J.H., 1979. Factor indeterminacy in the 1930's and the 1970's, some interesting parallels. Psychometrika 44, 157–167.
Stewart, G.W., 1980. The efficient generation of random orthogonal matrices with an application to condition estimators. SIAM J. Numer. Anal. 17, 403–409.
Ten Berge, J.M.F., 1998. Some recent developments in factor analysis and the search for proper communalities. In: Rizzi, A., Vichi, M., Bock, H.H. (Eds.), Advances in Data Science and Classification. Springer, Heidelberg, Germany.
Ten Berge, J.M.F., Kiers, H.A.L., 1991. A numerical approach to the approximate and the exact minimum rank of a covariance matrix. Psychometrika 56, 309–315.
Ten Berge, J.M.F., Nevels, K., 1977. A general solution to Mosier's oblique Procrustes problem. Psychometrika 42, 593–600.
Thurstone, L.L., 1935. The Vectors of Mind. University of Chicago Press, Chicago.
Trendafilov, N.T., Unkel, S., 2011. Exploratory factor analysis of data matrices with more variables than observations. J. Comput. Graph. Statist. 20, 874–891.
Unkel, S., Trendafilov, N.T., 2010. Simultaneous parameter estimation in exploratory factor analysis: an expository review. Internat. Statist. Rev. 78, 363–382.
Unkel, S., Trendafilov, N.T., 2013. Zig-zag exploratory factor analysis with more variables than observations. Comput. Statist. 28, 107–125.
Wechsler, D., 1991. WISC-III Manual. Psychological Corporation, San Antonio, TX.
Wilson, E., 1928. On hierarchical correlation systems. Proc. Natl. Acad. Sci. 14, 283–291.
Zhang, G., 2014. Estimating standard errors in exploratory factor analysis. Multivariate Behav. Res. 49, 339–353.