A Fixed Factor Analysis Procedure as an Extension of Principal Component Analysis

Kohei Adachi
Department of Human Sciences, Osaka University
1-2 Yamadaoka, Suita
Osaka 565-0871, Japan
E-mail: adachi@hus.osaka-u.ac.jp

1. Introduction

The prevailing procedures for exploratory factor analysis are underlain by the random factor model (Anderson & Rubin, 1956) in which factor scores are treated as random latent variables. Using $\mathbf{x}$ for a $p \times 1$ observation vector whose expectation E($\mathbf{x}$) equals the $p \times 1$ zero vector $\mathbf{0}_p$, the model is written as

$$\mathbf{x} = \mathbf{\Lambda f} + \mathbf{\Psi u} \, . \tag{1}$$

Here, $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ are unknown fixed parameter matrices, while $\mathbf{f}$ and $\mathbf{u}$ are random vectors: $\mathbf{\Lambda}$ is a p-variables × m-factors loading matrix and $\mathbf{\Psi}$ is a $p \times p$ diagonal matrix, whereas $\mathbf{f}$ ($m \times 1$) contains common factor scores, and $\mathbf{u}$ ($p \times 1$) is filled with standardised errors. The elements of $\mathbf{u}$ and the diagonal ones of $\mathbf{\Psi}^2$ are called unique factor scores and unique variances, respectively. It is assumed that E($\mathbf{f}$) = $\mathbf{0}_m$, E($\mathbf{ff}'$) = $\mathbf{I}_m$, E($\mathbf{u}$) = $\mathbf{0}_p$, E($\mathbf{uu}'$) = $\mathbf{I}_p$, and E($\mathbf{fu}'$) = ${}_m\mathbf{O}_p$, with $\mathbf{I}_m$ the $m \times m$ identity matrix and ${}_m\mathbf{O}_p$ the $m \times p$ matrix of zeros. The covariance matrix of $\mathbf{x}$ is then expressed as E($\mathbf{xx}'$) = $\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}^2$, which is fitted to its sample counterpart $\mathbf{S} = (n-1)^{-1}\mathbf{X}'\mathbf{X}$, where $\mathbf{X}$ is the $n \times p$ data matrix filled with the realizations of $\mathbf{x}'$ obtained from n individuals and is column-centered as $\mathbf{1}_n'\mathbf{X} = \mathbf{0}_p'$ with $\mathbf{1}_n$ the $n \times 1$ vector of ones. Popular procedures for fitting E($\mathbf{xx}'$) to $\mathbf{S}$ are unweighted least squares (ULS), weighted least squares (WLS), and maximum likelihood (ML) procedures, in which

$$h_{\mathrm{ULS}}(\mathbf{\Lambda},\mathbf{\Psi}^2 \,|\, \mathbf{S}) = \|\mathbf{S} - \mathbf{\Lambda\Lambda}' - \mathbf{\Psi}^2\|^2 = \mathrm{tr}(\mathbf{S} - \mathbf{\Lambda\Lambda}' - \mathbf{\Psi}^2)^2 \tag{2}$$

$$h_{\mathrm{WLS}}(\mathbf{\Lambda},\mathbf{\Psi}^2 \,|\, \mathbf{S}) = \mathrm{tr}\{(\mathbf{S} - \mathbf{\Lambda\Lambda}' - \mathbf{\Psi}^2)\mathbf{S}^{-1}\}, \tag{3}$$

$$h_{\mathrm{ML}}(\mathbf{\Lambda},\mathbf{\Psi}^2 \,|\, \mathbf{S}) = \log|\mathbf{\Lambda\Lambda}'+\mathbf{\Psi}^2|+\mathrm{tr}\{r\mathbf{S}(\mathbf{\Lambda\Lambda}'+\mathbf{\Psi}^2)^{-1}\}, \tag{4}$$

are, respectively, minimised over $\mathbf{\Lambda}$ and $\mathbf{\Psi}$. Here, (4) is proportional to the negative log likelihood derived under the multivariate normality assumption for $\mathbf{f}$ and $\mathbf{u}$, with $r = (n-1)/n$. I refer to the minimisation of (2), (3), or (4) as a random factor analysis (FA) procedure in this paper.

A factor analysis procedure very different from the random FA has recently been presented by de Leeuw (2004) and Henk, A. L. Kiers in the University of Groningen and been elaborated by Unkel and Trendafilov (2010a), where the presentation by Kiers has been described in Sočan (2003). In this procedure, common and unique factor scores are treated rather as unknown fixed parameters and its loss function is expressed as

$$f_{\mathrm{ULS}}(\mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Psi} \,|\, \mathbf{X}) = \|\mathbf{X} - \mathbf{F\Lambda}' - \mathbf{U\Psi}\|^2 = \|\mathbf{X} - \mathbf{ZA}'\|^2 \, , \tag{5}$$

where $\mathbf{F}$ is an n-individuals × m-factors matrix of common factor scores, $\mathbf{U}$ ($n \times p$) contains unique factor scores, $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ is the $n \times r$ matrix containing both factor scores, and $\mathbf{A}$ ($p \times r$) = [$\mathbf{\Lambda}$,$\mathbf{\Psi}$] with $r = p + m$. Function (5) is minimised over $\mathbf{Z}$, $\mathbf{\Lambda}$, and $\mathbf{\Psi}$ subject to

$$\mathbf{1}_n'\mathbf{Z} = \mathbf{0}_q' \, , \tag{6}$$

$$\frac{1}{n}\mathbf{Z}'\mathbf{Z} = \mathbf{I}_q \ . \tag{7}$$

It should be noted that the constraints (6) and (7) summarize the conditions $\mathbf{1}_n'\mathbf{F} = \mathbf{0}_m'$, $n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_m$, $\mathbf{1}_n'\mathbf{U} = \mathbf{0}_p'$, $\mathbf{U}'\mathbf{U} = \mathbf{I}_p$, and $n^{-1}\mathbf{F}'\mathbf{U} = {}_m\mathbf{O}_p$ which correspond to the assumptions for **f** and **u** in the random FA. I refer to the above minimisation as fixed FA, since factor scores are treated as fixed parameters to be estimated. A feature of the fixed FA is to estimate $\mathbf{\Psi}$, $\mathbf{\Lambda}$ and factor score matrix **Z** simultaneously. On the other hand, in the random FA where only $\mathbf{\Psi}$ and $\mathbf{\Lambda}$ are estimated, we must resort to a two stage procedure if **F** and **U** are needed to be obtained: they are obtained using the resulting $\mathbf{\Psi}$ and $\mathbf{\Lambda}$ in a post hoc manner. The fixed FA can be viewed as an extension of the principal component analysis (PCA) formulated as minimizing $\|\mathbf{X} - \mathbf{F}\mathbf{\Lambda}'\|^2$ over **F** and $\mathbf{\Lambda}$: this function is extended as (5) by introducing $\mathbf{U}\mathbf{\Psi}$ in the fixed FA.

The algorithm for the fixed FA consists of alternately iterating the following three steps; [A] the minimisation of (5) over **Z** subject to (6) and (7) while keeping $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ fixed, [B] minimising (5) over $\mathbf{\Lambda}$ with **Z** and $\mathbf{\Psi}$ kept fixed, and [C] the minimisation of (5) over the diagonal elements of $\mathbf{\Psi}$ with **Z** and $\mathbf{\Lambda}$ fixed. The algorithm is simple in that the problem in each step can be solved explicitly: the solutions in [A], [B], and [C] are given by

$$\mathbf{Z} = [\mathbf{F}, \mathbf{U}] = \sqrt{n}\,\tilde{\mathbf{K}}\tilde{\mathbf{L}}' , \tag{8}$$

$$\mathbf{\Lambda} = \frac{1}{n}\mathbf{X}'\mathbf{F} , \tag{9}$$

$$\mathbf{\Psi} = \frac{1}{n}\mathrm{diag}(\mathbf{X}'\mathbf{U}) , \tag{10}$$

where $\tilde{\mathbf{K}}$ and $\tilde{\mathbf{L}}$ follow from the singular value decomposition (SVD) of **XA** defined as $\mathbf{XA} = \tilde{\mathbf{K}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{L}}'$ with $\tilde{\mathbf{K}}$ and $\tilde{\mathbf{L}}$ being column-orthonormal and $\tilde{\mathbf{\Lambda}}$ a $q \times q$ diagonal matrix and (de Leeuw, 2004; Sočan, 2003; Unkel & Trendafilov, 2010a). Though **Z** in (8) cannot be uniquely determined since the rank of **XA** does not exceed $p < q$ and the $q - p$ diagonal elements of $\tilde{\mathbf{\Lambda}}$ are zeros, the algorithm allows (5) to converge to a minima by setting **Z** at a matrix in the set of the optimal **Z**s.

Unkel and Trendafilov (2010a) have described that the fixed FA is a promising approach in the following points. First, in the former, all model unknowns can be simultaneously estimated. Second, $\mathbf{\Psi}$ is estimated in the fixed FA and thus the improper solutions are never given in which the estimates of unique variances (i.e., the diagonal elements of $\mathbf{\Psi}^2$) include a negative value, whereas such solutions are encountered in the random FA with $\mathbf{\Psi}^2$ estimated. Third, the algorithm for the fixed FA facilitates the application of exploratory factor analysis, because, as well as PCA, it is based on computationally well-known and efficient numerical procedure of the SVD of data matrices. Unkel and Trendafilov (2010b) further have utilized the merits of the fixed FA to develop a its robust version. However, some problems are thought to remain to be studied in the fixed FA. Among them, the problems concerning [1] scale invariance, [2] the matrix necessary for analyses, [3] the estimation of factor scores, and [4] exactness in parameter estimation are focused on in this paper.

The first problem is that the loss function (5) for the fixed FA is not scale invariant with $\|\mathbf{X} - \mathbf{F}\mathbf{\Lambda}' - \mathbf{U}\mathbf{\Psi}\|^2 \neq \|\mathbf{X}\mathbf{D} - \mathbf{F}\mathbf{\Lambda}'\mathbf{D} - \mathbf{U}\mathbf{\Psi}\mathbf{D}\|^2$ for a diagonal matrix $\mathbf{D} \neq \mathbf{I}_p$; an original data set and its standardised version lead to different solutions. Though ULS function (2) is also not scale invariant, the WLS and ML loss functions (3) and (4) in the random FA are scale invariant with $h_{\mathrm{WLS}}(\mathbf{\Lambda},\mathbf{\Psi}^2 \,|\, \mathbf{S}) = h_{\mathrm{WLS}}(\mathbf{D\Lambda}, \mathbf{D\Psi}^2\mathbf{D} \,|\, \mathbf{DSD})$ and $h_{\mathrm{ML}}(\mathbf{\Lambda},\mathbf{\Psi}^2 \,|\, \mathbf{S}) = h_{\mathrm{ML}}(\mathbf{D\Lambda}, \mathbf{D\Psi}^2\mathbf{D} \,|\, \mathbf{DSD})$ (e.g., Yanai & Ichikawa, 2007). In this paper, I extend (5) into a weighted least squares (WLS) function which has the scale invariant property if the weights are chosen suitably and also includes (5) as a special case. I refer to the approach with the WLS function as a WLS fixed FA procedure and the minimisation of (5) as the ULS fixed FA procedure.

Second, it *seems* to be a drawback that the fixed FA apparently *seems* to need the original data matrix **X**

of a larger size than its covariance matrix **S**, which suffices for estimating $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ in the random FA. In this paper, however, the fact is shown that only a covariance matrix is sufficient for minimising loss function (5) and its WLS version to be presented. As seen in Section 2.2, this fact is underlain by that the covariance matrix between variables and factors is uniquely determined.

Third, the factor score matrix $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ cannot be uniquely determined in the ULS fixed FA and also in its WLS version. We must thus choose **Z** among multiple candidates of its estimators, but how to choose it has not been mentioned in the previous literatures (de Leeuw, 2004; Sočan, 2003; Unkel & Trendafilov, 2010). I thus propose a choice of **Z**. The fourth problem is to answer the question whether the parameter matrices ($\mathbf{\Lambda}$ and $\mathbf{\Psi}$) are estimated more exactly in the fixed or random FA. Further, factor scores $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ can also be obtained using two stage procedures in the random FA and it is thus included in the problem to compare the FA procedures in the exactness of factor score estimation.

In the next section, I detail the WLS fixed FA procedure with answering the first, second and third problems, which is followed by the Section 3 where I report the simulation study concerning the fourth problem.

2. Weighted least squares fixed factor analysis

The loss function to be minimised in the proposed WLS fixed FA is introduced in Section 2.1. Properties of the factor score matrix $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ that minimises the WLS function are discussed in Section 2.2, which leads to the update formula of $\mathbf{A} = [\mathbf{\Lambda}, \mathbf{\Psi}]$ in Section 2.3. There, we can find that only a sample covariance matrix suffices for updating **A** and minimizing the loss function. The iterative algorithm for the minimization is summarized in Section 2.4, and an estimator of **Z** is presented in Section 2.5.

2.1. Loss function

I propose a WLS fixed FA procedure, in which a WLS loss function

$$f(\mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Psi} \,|\, \mathbf{X}) = \mathrm{tr}(\mathbf{X} - \mathbf{Z}\mathbf{A}')\mathbf{W}^{-1}(\mathbf{X} - \mathbf{Z}\mathbf{A}')' = \|\mathbf{X} - \mathbf{Z}\mathbf{A}'\|^2_{\mathbf{W}^{-1}} = \|\mathbf{X} - \mathbf{F}\mathbf{\Lambda}' - \mathbf{U}\mathbf{\Psi}\|^2_{\mathbf{W}^{-1}}, \tag{11}$$

is minimised over $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ and $\mathbf{A} = [\mathbf{\Lambda}, \mathbf{\Psi}]$ subject to conditions (6) and (7). For matrix **W**, I choose

$$\mathbf{W} = \mathbf{S} = \frac{1}{n-1}\mathbf{X}'\mathbf{X} \quad . \tag{12}$$

Then, function (11) is scale invariant with $f(\mathbf{Z}, \mathbf{D\Lambda}, \mathbf{D\Psi} \,|\, \mathbf{XD}) = n\mathrm{tr}(\mathbf{XD} - \mathbf{ZA}'\mathbf{D})[(n-1)^{-1}\mathbf{DX}'\mathbf{XD}]^{-1}(\mathbf{XD} - \mathbf{ZA}'\mathbf{D})' = \mathrm{tr}(\mathbf{X} - \mathbf{ZA}')\mathbf{S}^{-1}(\mathbf{X} - \mathbf{ZA}')' = f(\mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Psi} \,|\, \mathbf{X})$. However, I continue to use the symbol **W** (not **S**) to allow the following discussions to include the case of the ULS approach with **W** set at $\mathbf{I}_p$.

2.2. Properties of Optimal Factor Scores

Let us consider the optimal factor score matrix $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ that minimises (11) subject to (6) and (7) for given $\mathbf{A} = [\mathbf{\Lambda}, \mathbf{\Psi}]$. Using c for the part irrelevant to **Z** in (11), it is rewritten as $c - \mathrm{tr}\mathbf{X}'\mathbf{ZA}'\mathbf{W}^{-1}$. This minimisation thus amounts to maximizing $g(\mathbf{Z}) = \mathrm{tr}\mathbf{X}'\mathbf{ZA}'\mathbf{W}^{-1} = \mathrm{tr}\mathbf{Z}'\mathbf{XW}^{-1}\mathbf{A}$ over **Z**. The maximization under (6) and (7) is achieved using the SVD of $n^{-1/2}\mathbf{XW}^{-1}\mathbf{A}$ $(n \times q)$ defined as

$$\frac{1}{\sqrt{n}}\mathbf{XW}^{-1}\mathbf{A} = \mathbf{K\Theta L}' = [\mathbf{K}_1, \mathbf{K}_2]\begin{bmatrix}\mathbf{\Theta}_1 & \\ & {}_m\mathbf{O}_m\end{bmatrix}\begin{bmatrix}\mathbf{L}_1' \\ \mathbf{L}_2'\end{bmatrix} = \mathbf{K}_1\mathbf{\Theta}_1\mathbf{L}_1' \,. \tag{13}$$

Here, $\mathrm{rank}(\mathbf{XW}^{-1}\mathbf{A}) = p < q$, $\mathbf{\Theta}_1$ is the $p \times p$ positive definite diagonal matrix, $\mathbf{\Theta}$ is the $q \times q$ diagonal matrix whose diagonal blocks are $\mathbf{\Theta}_1$ and ${}_m\mathbf{O}_m$, and $\mathbf{K} = [\mathbf{K}_1, \mathbf{K}_2]$ $(n \times q)$ and $\mathbf{L} = [\mathbf{L}_1, \mathbf{L}_2]$ $(q \times q)$ satisfy

$$\mathbf{K}'\mathbf{K} = \mathbf{L}'\mathbf{L} = \begin{bmatrix}\mathbf{K}_1' \\ \mathbf{K}_2'\end{bmatrix}[\mathbf{K}_1, \mathbf{K}_2] = \begin{bmatrix}\mathbf{L}_1' \\ \mathbf{L}_2'\end{bmatrix}[\mathbf{L}_1, \mathbf{L}_2] = \mathbf{I}_q \,, \tag{14}$$

with $\mathbf{K}_1$, $\mathbf{K}_2$, $\mathbf{L}_1$, and $\mathbf{L}_2$ being $n \times p$, $n \times m$, $q \times p$, and $q \times m$ matrices, respectively. We have the inequality

$$g(\mathbf{Z}) = \mathrm{tr}\mathbf{Z}'\mathbf{X}\mathbf{W}^{-1}\mathbf{A} = \mathrm{tr}\, n^{1/2}\mathbf{Z}'\mathbf{K}\Theta\mathbf{L}' = \mathrm{tr}\, n\mathbf{L}'(n^{-1/2}\mathbf{Z})'\mathbf{K}\Theta \le \mathrm{tr}\Theta = \mathrm{tr}\Theta_1 . \tag{15}$$

(ten Berge, 1983), and the upper bound $\mathrm{tr}\Theta_1$ is attained for

$$\mathbf{Z} = \sqrt{n}\,\mathbf{K}\mathbf{L}' = \sqrt{n}\,\mathbf{K}_1\mathbf{L}_1' + \sqrt{n}\,\mathbf{K}_2\mathbf{L}_2' \tag{16}$$

under (7). Obviously, $\mathbf{K}_2\mathbf{L}_2'$ and $\mathbf{Z}$ cannot be uniquely determined. These facts with (13) - (16) for the case with $\mathbf{W} = \mathbf{I}_p$ have been shown in de Leeuw, 2004, Sočan, 2003, and Unkel and Trendafilov (2010a) .

Although in those literatures it has not been described that (16) satisfies not only (7) but also the other condition (6), this fact is shown as follows. Since (13) implies $n^{1/2}\mathbf{K}_1 = \mathbf{X}\mathbf{W}^{-1}\mathbf{A}\mathbf{L}_1\Theta_1^{-1}$, the first term in the right-hand side of (16) can be expressed as

$$\sqrt{n}\,\mathbf{K}_1\mathbf{L}_1' = \mathbf{X}\mathbf{W}^{-1}\mathbf{A}\mathbf{L}_1\Theta_1^{-1}\mathbf{L}_1' , \tag{17}$$

which implies $\mathbf{1}_n'\mathbf{K}_1 = \mathbf{0}_p$. The non-unique second term $n^{1/2}\mathbf{K}_2\mathbf{L}_2'$ can be chosen such that $\mathbf{1}_n'\mathbf{K}_2 = \mathbf{0}_m$ as seen in Appendix. Using $\mathbf{1}_n'\mathbf{K}_1 = \mathbf{0}_p'$ and $\mathbf{1}_n'\mathbf{K}_2 = \mathbf{0}_m'$ in (16), we can find that it satisfies condition (6).

It is an important fact for our proposed algorithm that, though $\mathbf{Z}$ is not unique, a p-variables × q-factors covariance matrix $n^{-1}\mathbf{X}'\mathbf{Z}$ is uniquely determined as

$$\frac{1}{n}\mathbf{X}'\mathbf{Z} = \frac{1}{n}\mathbf{X}'[\mathbf{F}, \mathbf{U}] = \mathbf{W}'\mathbf{A}'^{+}\mathbf{L}_1\Theta_1\mathbf{L}_1' , \tag{18}$$

if Θ_1 has distinct diagonal elements and $\mathbf{A}$ ($p \times q$) is of full-row rank which implies $\mathbf{A}\mathbf{A}^{+} = \mathbf{I}_p$. This fact is proved as follows. Using $\mathbf{A}\mathbf{A}^{+} = \mathbf{I}_p$ and (13) we have $n^{-1/2}\mathbf{X} = n^{-1/2}\mathbf{X}\mathbf{W}^{-1}\mathbf{A}\mathbf{A}^{+}\mathbf{W} = \mathbf{K}_1\Theta_1\mathbf{L}_1'\mathbf{A}^{+}\mathbf{W}$. This equation and (16) imply that $n^{-1}\mathbf{X}'\mathbf{Z} = (n^{-1/2}\mathbf{X})'(n^{-1/2}\mathbf{Z}) = (\mathbf{W}'\mathbf{A}'^{+}\mathbf{L}_1\Theta_1\mathbf{K}_1')(\mathbf{K}\mathbf{L}')$ which leads to (18).

As seen in the next section, $\mathbf{A}$ can be updated using (18), and the $\mathbf{L}_1$ in (18) can be obtained from the eigenvalue decomposition (EVD) defined as

$$r\mathbf{A}'\mathbf{W}^{-1}\mathbf{S}\mathbf{W}^{-1}\mathbf{A} = \mathbf{L}_1\Theta_1{}^2\mathbf{L}_1' , \tag{19}$$

even if we do not have the original data matrix $\mathbf{X}$ but only the covariance matrix $\mathbf{S}$.

2.3. Update of Loadings and Unique Variances

First, let us consider minimising function (11) over Λ while keeping $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ and Ψ fixed. Using the conditions $n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_m$ and $n^{-1}\mathbf{F}'\mathbf{U} = {}_m\mathbf{O}_p$ implied by (7), function (11) is rewritten as $c^{\#} - 2\mathrm{tr}\mathbf{X}'\mathbf{F}\Lambda'\mathbf{W}^{-1} + n\,\mathrm{tr}\Lambda\Lambda'\mathbf{W}^{-1}$ with $c^{\#}$ the constant irrelevant to Λ. We can thus find that (11) is minimized when Λ equals the matrix $n^{-1}\mathbf{X}'\mathbf{F}$, i.e., (9), which consists of the first p columns of (18) and does not depend on Ψ. Next, let us consider minimising (11) over diagonal matrix Ψ with $\mathbf{Z}$ and Λ kept fixed. Using $n^{-1}\mathbf{U}'\mathbf{U} = \mathbf{I}_p$ and $n^{-1}\mathbf{F}'\mathbf{U} = {}_m\mathbf{O}_p$, (11) is rewritten as $c^{*} - 2\mathrm{tr}[\mathrm{diag}(\mathbf{W}^{-1}\mathbf{X}'\mathbf{U})\Psi^2] + n\,\mathrm{tr}[\mathrm{diag}(\mathbf{W}^{-1})\Psi^2]$ with c^{*} expressing the part irrelevant to Ψ and $\mathrm{diag}(\mathbf{W}^{-1})$ denoting the diagonal matrix whose diagonal elements are those of $\mathbf{W}^{-1}$. We can thus find that the minimisation is attained for $\Psi = \mathrm{diag}(\mathbf{W}^{-1})^{-1}\mathrm{diag}(\mathbf{W}^{-1}\mathbf{V})$, where $\mathbf{V} = n^{-1}\mathbf{X}'\mathbf{U}$ consists of the last m columns of (16) and this Ψ does not depend on Λ.

The above two results show that the update of $\mathbf{A} = [\Lambda, \Psi]$ by

$$\Lambda = \frac{1}{n}\mathbf{X}'\mathbf{F} = \mathbf{W}'\mathbf{A}_{\mathrm{OLD}}{}^{+\prime}\mathbf{L}_1\Theta_1\mathbf{L}_1'\mathbf{H}^{[m]} , \tag{20}$$

$$\Psi = \mathrm{diag}\,(\mathbf{W}^{-1})^{-1}\mathrm{diag}(\mathbf{W}^{-1}\frac{1}{n}\mathbf{X}'\mathbf{U}) = \mathrm{diag}\,(\mathbf{W}^{-1})^{-1}\mathrm{diag}(\mathbf{W}^{-1}[\mathbf{W}'\mathbf{A}_{\mathrm{OLD}}{}^{+\prime}\mathbf{L}_1\Theta_1\mathbf{L}_1']\,\mathbf{H}_p) \tag{21}$$

minimises (11) for given $\mathbf{Z}$, where $\mathbf{H}^{[m]} = [\mathbf{I}_m, {}_m\mathbf{O}_p]'$ ($q \times m$), $\mathbf{H}_p = [{}_p\mathbf{O}_m, \mathbf{I}_p,]'$ ($q \times p$), and $\mathbf{A}_{\mathrm{OLD}}$ denotes the matrix $\mathbf{A}$ before the update.

The resulting Λ and Ψ allow us to express the loss function value as

$$f(\mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Psi}) = n \operatorname{tr}\{r\,\mathbf{S} - (\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}^2)\}\mathbf{W}^{-1}\,, \tag{22}$$

which is derived by the use of (20) and (21) in the function (11) expanded as $\operatorname{tr}(\mathbf{X}'\mathbf{X}+n\mathbf{\Lambda}\mathbf{\Lambda}'+n\mathbf{\Psi}^2 -2\mathbf{X}'\mathbf{F}\mathbf{\Lambda}'-2\mathbf{X}'\mathbf{U}\mathbf{\Psi})\mathbf{W}^{-1}$. The function value in (22) is standardized into

$$f_s(\mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Psi}) = \frac{f(\mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Psi})}{nr \operatorname{tr}\mathbf{S}\mathbf{W}^{-1}}\,, \tag{23}$$

which takes a value of zero to one.

2.4. Iterative Algorithm

Given $\mathbf{S}$, loss function (11) can be minimised subject to (6) and (7) simply by iterating the update with (20) and (21), without updating $\mathbf{Z}$. The iterative algorithm for the minimisation thus follows the next steps:

Step 1. Initialize $\mathbf{A} = [\mathbf{\Lambda}, \mathbf{\Psi}]$.
Step 2. Perform EVD in (19) and obtain (18).
Step 3. Update $\mathbf{A} = [\mathbf{\Lambda}, \mathbf{\Psi}]$ with (20) and (21).
Step 4. Finish if the decrease in (23) from the previous step is less than a constant ε;
otherwise, go back to Step 2.

In this paper, $\varepsilon = 0.1^7$ is used, and the EVD of the covariance matrix, i.e., $\mathbf{S} = \mathbf{B}\mathbf{\Phi}^2\mathbf{B}'$, is used for initializing $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ as $\mathbf{B}_m\mathbf{\Phi}_m$ and $\operatorname{diag}(\mathbf{S} - \mathbf{B}_m\mathbf{\Phi}_m{}^2\mathbf{B}_m')^{1/2}$, respectively, where the diagonal elements of diagonal matrix $\mathbf{\Phi}$ is arranged in descending order, $\mathbf{B}'\mathbf{B} = \mathbf{I}_p$, $\mathbf{B}_m$ containing the first m columns of $\mathbf{B}$, and $\mathbf{\Phi}_m$ the first $m \times m$ diagonal block of $\mathbf{\Phi}$.

2.5. Factor Score Estimator

We can obtain factor score matrix $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ after the steps in Section 2.4. However, the optimal $\mathbf{Z}$ expresses as (16), i.e., $\mathbf{Z} = n^{-1/2}\mathbf{K}_1\mathbf{L}_1' + n^{-1/2}\mathbf{K}_2\mathbf{L}_2'$ is not unique: though the first term $n^{-1/2}\mathbf{K}_1\mathbf{L}_1'$ is unique, the second $n^{-1/2}\mathbf{K}_2\mathbf{L}_2'$ is the product of any $\mathbf{K}_2$ and $\mathbf{L}_2$ satisfying (14). Though it might be a strategy to impose a constraint onto $\mathbf{K}_2\mathbf{L}_2'$ such that it is uniquely determined, a reasonable constraint is difficult to find. I thus choose the term

$$\hat{\mathbf{Z}} = \sqrt{n}\,\mathbf{K}_1\mathbf{L}_1' \tag{24}$$

determined uniquely in (16) as the estimator of $\mathbf{Z}$. This has one drawback, but two kinds of optimality.

The drawback is that (24) does not satisfy (7) with $n^{-1}\hat{\mathbf{Z}}'\hat{\mathbf{Z}} = \mathbf{L}_1\mathbf{L}_1' \neq \mathbf{I}_q$ since the rank of $\hat{\mathbf{Z}}$ is $p < q$, though (24) satisfies (6). One of the two optimal properties is that (24) attains the upper bound of (15) with $g(\hat{\mathbf{Z}}) = \operatorname{tr}\hat{\mathbf{Z}}'\mathbf{X}\mathbf{W}^{-1}\mathbf{A} = \operatorname{tr}\mathbf{L}_1\mathbf{K}_1'\mathbf{K}\mathbf{\Theta}\mathbf{L}' = \operatorname{tr}\mathbf{\Theta}_1$. The other property is that (24) is the reduced rank approximation of the factor score matrix $\mathbf{Z}$ of rank q; the $n \times q$ matrix $\mathbf{Y}$ of rank $p < q$ that minimises

$$\eta(\mathbf{Y}) = \|\mathbf{Y} - \mathbf{Z}\|^2 = \|\mathbf{Y} - \sqrt{n}\,\mathbf{K}\mathbf{L}'\|^2 \tag{25}$$

is given by (24) with $\eta(\mathbf{Y}) \geq \eta(\hat{\mathbf{Z}})$ for any $n \times q$ matrix $\mathbf{Y}$.

3. Simulation study

I performed a simulation study to compare random and fixed FA procedures in how well true loadings, unique variances, and factor scores can be recovered.

3.1. Procedures

The underlying models are different between the random FA and the fixed FA; factor scores are treated as random variables in the former, but as fixed parameters in the latter. However, comparisons of both types

of FA are difficult unless data sets are synthesized with the same model. I thus choose to use the random factor model (1), as it seems to be reasonable that a new approach (the fixed FA) is tested by a prevailing one (the model for the random FA), as compared to testing a prevailing approach by a new one, although this choice is thought disadvantageous for the fixed FA.

A thousand data sets satisfying (1) were generated by replicating the following procedures: [1] The sizes of matrices were randomly chosen: I drew m from $DU(1, 5)$, p from $DU(4m, 8m)$, and then n from $DU(8p, 12p)$, with $DU(a, b)$ denoting the discrete uniform distribution of the integers from a to b. [2] $\mathbf{\Lambda}$ was filled with $u_{[-1.0,\ 1.0]}$ and the diagonal elements of $\mathbf{\Psi}$ were filled with $u_{[\sqrt{0.1},\sqrt{0.7}]}$, where $u_{[\alpha,\ \beta]}$ denotes a variable distributed uniformly over the range from α and β; [3] The rows of data matrix $\mathbf{X}$ were filled with the $\mathbf{x}'$ satisfying (1), i.e., $\mathbf{x} = \mathbf{\Lambda f} + \mathbf{\Psi u}$, where $\mathbf{f}$ was drawn from $N_m(\mathbf{0}_m, \mathbf{I}_m)$ and $\mathbf{u}$ was drawn from $N_p(\mathbf{0}_p, \mathbf{I}_p)$ independently of $\mathbf{f}$, with $N_m(\mathbf{0}_m, \mathbf{I}_m)$ denoting the m-variate normal distribution whose mean vector and covariance matrix are $\mathbf{0}_m$ and $\mathbf{I}_m$. Further, the n sets of $\mathbf{f}'$ and those of $\mathbf{u}'$ constituted the rows of $\mathbf{F}$ and $\mathbf{U}$, respectively.

The reason for randomly choosing matrix sizes in the above step [1] was to cover widely usual factor analysis cases where the number of n, p, and m are proportional to each other with n much larger than p and $p > m$. Each data matrix $\mathbf{X}$ was standardized so that column averages and variances were zeros and ones, respectively, to be matched to the usual application cases. Thus, the loading matrix and the diagonal matrix including unique variances to be recovered were $\mathrm{diag}(\mathbf{\Sigma})^{-1/2}\mathbf{\Lambda}$ and $\mathrm{diag}(\mathbf{\Sigma})^{-1}\mathbf{\Psi}^2$, respectively, with $\mathbf{\Sigma} = \mathbf{\Lambda\Lambda}' + \mathbf{\Psi}^2$. I use just $\mathbf{X}$, $\mathbf{\Lambda}$ and $\mathbf{\Psi}^2$ for the standardised $\mathbf{X}$, $\mathrm{diag}(\mathbf{\Sigma})^{-1/2}\mathbf{\Lambda}$, and $\mathrm{diag}(\mathbf{\Sigma})^{-1}\mathbf{\Psi}^2$, respectively.

Each $\mathbf{X}$ was analyzed by the following five procedures; the ULS random FA in which Harman and Jones' (1966) MINRES algorithm was used for minimising (2), the WLS random FA in which Lee's (1978) algorithm was used for minimizing (3), the ML random FA in which Rubin and Thayer's (1982) EM algorithm was used for minimising (4), the ULS fixed FA for minimising (5) or (11) with $\mathbf{W} = \mathbf{I}_p$, and the WLS fixed FA for minimising (11) with $\mathbf{W} = \mathbf{S}$.

In the random FA procedures, $\mathbf{\Lambda}$ and $\mathbf{\Psi}^2$ are initialized in the same manner as described in Section 2.4, and the iteration was stopped when the changes in $h_{\mathrm{ULS}}(\mathbf{\Lambda},\mathbf{\Psi}^2 \,|\, \mathbf{S})/\mathrm{tr}\mathbf{S}^2$, $h_{\mathrm{WLS}}(\mathbf{\Lambda},\mathbf{\Psi}^2 \,|\, \mathbf{S})/p$, and $h_{\mathrm{ML}}(\mathbf{\Lambda},\mathbf{\Psi}^2 \,|\, \mathbf{S})/(|r\mathbf{S}|+p)$ became less than 0.1^7 in the ULS, WLS, and ML procedures. Using the resulting $\mathbf{\Lambda}$ and $\mathbf{\Psi}^2$, I obtained the regression estimate of common factor score matrix $\mathbf{F} = \mathbf{X\Psi}^{-2}\mathbf{\Lambda}(\mathbf{I}_m + \mathbf{\Lambda}'\mathbf{\Psi}^{-2}\mathbf{\Lambda})^{-1}$ which gave unique factor score matrix $\mathbf{U} = (\mathbf{X} - \mathbf{F\Lambda}')\mathbf{\Psi}^{-1}$ subsequently.

3.2. Results

The ULS and WLS random FA procedures yielded non-positive $\mathbf{\Psi}$s for 14 and 12 data sets, respectively, among 1000 ones, where the non-positive $\mathbf{\Psi}$ refers to the $\mathbf{\Psi}$ with at least one diagonal element equal to or less than 0.0. A reason for the ML random FA procedure not giving such a solution is that the data were completely matched to the model underling this procedure, that is, (1) with the normality assumption. Also, the fixed FA never gave a non-positive $\mathbf{\Psi}$, which shows an advantage of the fixed FA. The solutions with the non-positive $\mathbf{\Psi}$ were removed, that is, the 986 (=1000 −14) and 988 (=1000 −12) sets of solutions were considered for ULS and WLS random FA, respectively, in the following procedures.

As an index of the recovery of true matrices, I use a standardized similarity defined as

$$\mathrm{SS}(\hat{\mathbf{M}},\mathbf{M}) = 1 - \frac{0.5\left\|\hat{\mathbf{M}} - \mathbf{M}\right\|^2}{\left\|\hat{\mathbf{M}} - \overline{\mathbf{M}}\right\|^2 + \left\|\mathbf{M} - \overline{\mathbf{M}}\right\|^2}, \tag{26}$$

where $\hat{\mathbf{M}}$ $(N \times M)$ is the estimated counterpart of true matrix $\mathbf{M}$ and $\overline{\mathbf{M}}$ is filled with the average of the NM elements in $0.5(\hat{\mathbf{M}}+\mathbf{M})$ (Adachi, 2011). Index (26) takes a value of zero to one and attains the upper bound one for $\hat{\mathbf{M}} = \mathbf{M}$. Table 1 shows the quartiles of the standardized similarities for loadings, unique variances, and factor scores, i.e., $\mathrm{SS}(\hat{\mathbf{\Lambda}}, \mathbf{\Lambda})$, $\mathrm{SS}(\hat{\mathbf{\Psi}}^2\mathbf{1}_p, \mathbf{\Psi}^2\mathbf{1}_p)$, $\mathrm{SS}(\hat{\mathbf{F}}, \mathbf{F})$, and $\mathrm{SS}(\hat{\mathbf{U}}, \mathbf{U})$, in the solutions for each procedure. The quartiles of $\mathrm{CC}(\hat{\mathbf{\Lambda}}, \mathbf{\Lambda})$, $\mathrm{CC}(\hat{\mathbf{\Psi}}^2\mathbf{1}_p, \mathbf{\Psi}^2\mathbf{1}_p)$, $\mathrm{CC}(\hat{\mathbf{F}}, \mathbf{F})$, and $\mathrm{CC}(\hat{\mathbf{U}}, \mathbf{U})$ are also shown,

Table 1. The 1st, 2nd, and 3rd quartiles of the standardised similarities (SS) and correlation coefficients (CC) for $\mathbf{\Lambda}$, $\mathbf{\Psi}^2\mathbf{1}_p$, $\mathbf{F}$, and $\mathbf{U}$ resulting from random and fixed FA procedures.

In-dex	Fac-tor	Proce-dure	$\mathbf{\Lambda}$: Loadings			$\mathbf{\Psi}^2$: Unique Variances			$\mathbf{F}$: Common Factors			$\mathbf{U}$: Unique Factors		
			1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
SS	Ran-dom	ULS	0.996	0.998	0.998	0.968	0.982	0.989	0.970	0.981	0.987	0.941	0.951	0.959
		WLS	0.996	0.998	0.998	0.909	0.940	0.958	0.969	0.981	0.987	0.943	0.952	0.959
		ML	0.996	0.998	0.998	0.969	0.983	0.989	0.969	0.981	0.987	0.941	0.951	0.959
	Fixed	ULS	0.996	0.998	0.998	0.966	0.982	0.989	0.959	0.979	0.986	0.918	0.929	0.939
		WLS	0.996	0.998	0.998	0.943	0.964	0.977	0.960	0.980	0.987	0.918	0.929	0.939
CC	Ran-dom	ULS	0.993	0.996	0.997	0.991	0.994	0.996	0.942	0.964	0.974	0.888	0.906	0.921
		WLS	0.994	0.996	0.997	0.987	0.992	0.994	0.942	0.964	0.975	0.886	0.905	0.919
		ML	0.993	0.996	0.997	0.991	0.994	0.996	0.941	0.963	0.974	0.887	0.906	0.921
	Fixed	ULS	0.993	0.996	0.997	0.991	0.994	0.996	0.922	0.960	0.974	0.837	0.859	0.879
		WLS	0.993	0.996	0.997	0.990	0.994	0.996	0.922	0.960	0.974	0.838	0.860	0.880

with CC($\hat{\mathbf{\Lambda}}$, $\mathbf{\Lambda}$) denoting the product moment correlation coefficient between vec$\hat{\mathbf{\Lambda}}$ and vec$\mathbf{\Lambda}$, as the coefficient is familiar and thus easy to capture its largeness/smallness.

In Table 1 we find that the exactness in the recovery of loading matrix $\mathbf{\Lambda}$ was equivalent between five procedures and that for unique variance $\mathbf{\Psi}^2\mathbf{1}_p$ was also almost equivalent except for the WLS random FA with lower quartiles. The fixed FA procedures are found to be inferior to the random FA ones in the recovery of factor scores $\mathbf{F}$ and $\mathbf{U}$, which is a little surprising considering that the scores are estimated in the fixed FA, while they are obtained in a post-hoc manner in the random FA. It is conjectured that this result may be caused from the use of the data synthesizing procedure disadvantageous for the fixed FA or be due to undesirable properties of the estimator (24). However, the recovery of $\mathbf{F}$ and $\mathbf{U}$ by the fixed FA procedures is also thought satisfactory in that the quartiles of the standardised similarity exceeded 0.9 and those of correlation coefficient were sufficiently high.

4. Conclusions

In this paper, the fixed factor analysis (FA) was focused on in which the matrices of common and unique factor scores, loadings, and unique variances are treated as fixed parameters, and an weighted least squares (WLS) procedure for the fixed FA was presented which has the scale invariant property and includes the existing unweighted least squares (ULS) procedure as a special case. Further, in the fixed FA, as well as in the random FA, loadings and unique variances are shown to be estimated only with a sample covariance matrix, even if its original data matrix is not given. After the estimation, common and unique factor scores can be obtained using the data matrix, though the scores cannot be uniquely determined. It was proposed to use the reduced rank approximation of the optimal score matrix as its estimator.

The simulation study showed the WLS and ULS fixed FA procedures recover true loadings and unique variances as well as the prevailing random FA procedures, in spite of that the data were synthesized with the model for the random FA. This fact establishes that the new fixed FA can be included in a family of factor analysis procedures. However, the study also showed that the recovery of factor scores by the fixed FA, though satisfactory, is a little worse than the recovery by the random FA. This result suggests that factor score estimators different from reduced rank approximation (24) should be considered.

Finally, I must discuss that the fixed FA can also be viewed as a constrained principal component analysis (PCA), although the fixed FA was described as an extended PCA in the title and Section 1. That is, PCA can also be formulated as minimizing the function (5), i.e., $||\mathbf{X} - \mathbf{F}\mathbf{\Lambda}' - \mathbf{U}\mathbf{\Psi}||^2$, for the ULS fixed FA, over $\mathbf{F}$, $\mathbf{\Lambda}$, $\mathbf{U}$, and $\mathbf{\Psi}$ subject to the constraints (6) and (7) with $n^{-1}\mathbf{U}'\mathbf{U} = \mathbf{I}_p$ deleted. The solution of $\mathbf{F}\mathbf{\Lambda}'$ and $\mathbf{U}\mathbf{\Psi}$ is given by the SVD of $\mathbf{X}$ and the loss function value attains zero with the rank of the resulting $\mathbf{U}\mathbf{\Psi}$ being

$p - m$. In this formulation, the fixed FA can be viewed as a version of PCA in which constraint $n^{-1}\mathbf{U}'\mathbf{U} = \mathbf{I}_p$ is added.

Appendix

We can set $\mathbf{K}_2 = (\mathbf{I}_n - \mathbf{K}_1^{+\prime}\mathbf{K}_1')\mathbf{G}$, with $\mathbf{G}$ being a matrix of $n \times m$ satisfying $\mathbf{1}_n'\mathbf{G} = \mathbf{0}_m$ and $\mathbf{G}'(\mathbf{I} - \mathbf{K}_1^{+\prime}\mathbf{K}_1')\mathbf{G} = \mathbf{I}_m$. Then, $\mathbf{1}_n'\mathbf{K}_2 = \mathbf{1}_n'(\mathbf{I}_n - \mathbf{K}_1^{+\prime}\mathbf{K}_1')\mathbf{G} = \mathbf{1}_n'\mathbf{G} = \mathbf{0}_m$ follows from that $\mathbf{1}_n'\mathbf{K}_1 = \mathbf{0}_p$ implies $\mathbf{1}_n'\mathbf{K}_1^{+\prime} = \mathbf{0}_p$.

REFERENCES

Adachi, K. (2011). Constrained principal component analysis of standardized data for biplots with unit-length variables vector. *Advanced in Data Analysis and Classification*, *5*, 23-36.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, *vol. 5*. (pp. 111-150). Berkeley, CA: University of California Press.

de Leeuw, J. (2004). Least squares optimal scaling of partially observed linear systems. In K. van Montfort, J. Oud & A. Satorra (Eds.), *Recent developments of structural equation models: Theory and applications*. Pp. 121-134. Dordrecht: Kluwer Academic Publishers.

Harman, H. H. & Jones, W. H. (1966). Factor analysis by minimizing residuals (Minres). *Psychomerika*, *31*, 351-369.

Lee, S. Y. (1978). The Gauss-Newton algorithm for the weighted least squares factor analysis. *Journal of the Royal Statistical Society*: *Series D* (*The Statistician*), *27*, *103-114*.

Rubin, D. B. & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, *47*, 69-76.

Sočan, G. (2003). *The incremental value of minimum rank factor analysis*. PhD Thesis, University of Groningen: Groningen.

ten Berge, J. M. F. (1983). A generalization of Kristof's theorem on the trace of certain matrix products. *Psychometrika*, *48*, 519-523.

Unkel, S. & Trendafilov, N. T. (2010a). Simultaneous parameter estimation in exploratory factor analysis: An expository review. *International Statistical Review*, *78*, 363-382.

Unkel, S. & Trendafilov, N. T. (2010b). A majorization algorithm for simultaneous parameter estimation in robust exploratory factor analysis. *Computational Statistics and Data Analysis*, *54*, 3348-3358.

Yanai, H. & Ichikawa, M. (2007). Factor analysis. In C.R. Rao & S. Sinharay (Eds.) *Handbook of statistics vol. 26: Psychometrics* (pp. 257-296). Amsterdam: Elsevier.

ABSTRACT

In this paper, a weighted least squares procedure for fixed factor analysis was proposed for simultaneously estimating common and unique factor scores, loadings, and unique variances. The proposed method, when the weight is given by the inverse of a sample covariance matrix, is scale invariant and also includes the existing fixed factor analysis procedure for the simultaneous estimation as a special case. Further, it was shown that the least square loss function can be minimized by iterating the update of loadings and unique variances, only a sample covariance matrix given without a raw data set, in spite of that it is matched to the model part in the function. After the convergence of the iteration, the estimates of common and unique factor scores can be obtained using the raw data set. In a simulation studies using the data synthesized with the random factor model, the proposed fixed factor analysis procedure was shown to recover true loadings and unique variances as well as the prevailing random factor analysis procedures.