Computational Statistics and Data Analysis 54 (2010) 2990–2998

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A new approach for selecting the number of factors☆

Yin-Ping Chen [a], Hsin-Cheng Huang [b], I-Ping Tu [b,*]

[a] Institute of Statistics, National Central University, Chung-Li, Taiwan
[b] Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 21 November 2008
Received in revised form 1 October 2009
Accepted 2 October 2009
Available online 24 October 2009

ABSTRACT

In factor analysis, it is critical to determine the number of factors. A new approach to select the number of factors based on unbiased risk estimation is introduced. This approach utilizes a concept, called generalized degrees of freedom (GDF), originally proposed for model selection in regression. A data perturbation technique is applied for estimating GDF. Simulation experiments show that the proposed method performs better than some commonly used methods.

1. Introduction

Factor analysis is used in a wide variety of fields such as psychometrics, finance, and some physical sciences. Suppose that we are given an $n \times p$ data matrix $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)'$ with $n > p$, where $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are independent and identically distributed p-dimensional column vectors. We consider the following orthogonal factor model:

$$\boldsymbol{Y}_i = \boldsymbol{\alpha} + \boldsymbol{A}\boldsymbol{f}_i + \boldsymbol{\varepsilon}_i; \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\alpha} \equiv E(\boldsymbol{Y}_i)$, $\boldsymbol{A}$ is a $p \times q$ constant matrix of factor loadings, $\boldsymbol{f}_i \sim N(\boldsymbol{0}, \boldsymbol{I}_q)$ is a q-dimensional column vector of common factors with $q < p$, $\boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{\Gamma})$ is a p-dimensional error vector of specific factors, independent of $\boldsymbol{f}_i$, and $\boldsymbol{\Gamma}$, usually referred to as the uniqueness matrix, is a diagonal matrix with diagonal elements $\gamma_1, \ldots, \gamma_p$. Then $\boldsymbol{Y}_i \sim N(\boldsymbol{\alpha}, \boldsymbol{A}\boldsymbol{A}' + \boldsymbol{\Gamma})$.

Several methods can be applied to estimate $\boldsymbol{\alpha}$, $\boldsymbol{A}$ and $\boldsymbol{\Gamma}$ (see Rencher, 2002). For example, based on q factors, the maximum likelihood (ML) estimate of $\boldsymbol{\alpha}$ is $\hat{\boldsymbol{\alpha}}_q^{(ML)} = \bar{\boldsymbol{Y}} \equiv \frac{1}{n}\sum_{i=1}^n \boldsymbol{Y}_i$, and the ML estimates $\hat{\boldsymbol{A}}_q^{(ML)}$ and $\hat{\boldsymbol{\Gamma}}_q^{(ML)}$ of $\boldsymbol{A}$ and $\boldsymbol{\Gamma}$ are obtained by maximizing the log-likelihood function:

$$l(\boldsymbol{A}, \boldsymbol{\Gamma}) \equiv -\frac{n}{2}\left\{p\log(2\pi) + \log|\boldsymbol{A}\boldsymbol{A}' + \boldsymbol{\Gamma}| + \mathrm{tr}\big((\boldsymbol{A}\boldsymbol{A}' + \boldsymbol{\Gamma})^{-1}\boldsymbol{S}\big)\right\},$$

where $\boldsymbol{S} \equiv \frac{1}{n}\sum_{i=1}^n \big(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\big)\big(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\big)'$ and $\hat{\boldsymbol{A}}_q^{(ML)}$ can be uniquely determined after applying some factor rotation criterion, such as the orthogonally rotated Varimax criterion (Kaiser, 1958) and the obliquely rotated Promax criterion (Hendrickson and White, 1964). By Thomson (1934), the factor score vector $\boldsymbol{f}_i$ based on q factors can be estimated by

$$\hat{\boldsymbol{f}}_{q,i} = \hat{\boldsymbol{A}}_q'\boldsymbol{S}^{-1}\big(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\big); \quad i = 1, \ldots, n, \tag{2}$$

where $\hat{\boldsymbol{A}}_q$ is an estimate of $\boldsymbol{A}$ based on q factors. Alternatively, a generalized least squares method (Bartlett, 1937, 1938) has also been applied to estimate $\boldsymbol{f}_i$'s.

☆ The software package designed for this approach, including an R function and its description, is available as supplementary material together with the electronic version of the paper.

* Corresponding address: Institute of Statistical Science, Academia Sinica, 128 Academia Rd. Sec. 2, Taipei 115, Taiwan. Tel.: +886 2 27835611x626; fax: +886 2 2783 1523.

E-mail address: iping@stat.sinica.edu.tw (I.-P. Tu).

0167-9473/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2009.10.002

It is known that the number of factors q plays a crucial role in factor analysis, because different choices of q can sometimes lead to different conclusions. In this paper, we consider selecting among $\{1, \ldots, q_{\max}\}$ with $q_{\max}$ a pre-specified integer. Note that $\hat{q}_{\max}$ has to satisfy $\frac{1}{2}\big((p-q_{\max})^2-(p+q_{\max})\big)\geq 0$ to avoid the model identifiability problem (see e.g., Bartholomew and Knott, 1999). Suppose that we are given some estimates $\hat{\boldsymbol{\alpha}}_q$, $\hat{\boldsymbol{A}}_q$, $\hat{\boldsymbol{\Gamma}}_q$ and $\{\hat{\boldsymbol{f}}_{q,i}\}$ of $\boldsymbol{\alpha}$, $\boldsymbol{A}$, $\boldsymbol{\Gamma}$ and $\{\boldsymbol{f}_i\}$, for $q = 1, \ldots, q_{\max}$. Let $\boldsymbol{X}_i \equiv \boldsymbol{\alpha} + \boldsymbol{A}\boldsymbol{f}_i$ and $\hat{\boldsymbol{X}}_{q,i} \equiv \hat{\boldsymbol{\alpha}}_q + \hat{\boldsymbol{A}}_q\hat{\boldsymbol{f}}_{q,i}$, for $q = 1, \ldots, q_{\max}$ and $i = 1, \ldots, n$. Our goal is to find $q \in \{1, \ldots, q_{\max}\}$ such that $Q(q) \equiv E[L(q)]$ is minimized, where

$$L(q) \equiv \frac{1}{n}\sum_{i=1}^{n}\big(\hat{\boldsymbol{X}}_{q,i}-\boldsymbol{X}_i\big)'\big(\hat{\boldsymbol{X}}_{q,i}-\boldsymbol{X}_i\big). \tag{3}$$

Several methods have been proposed to select the number of factors. For example, the number of factors may be selected according to the proportion of variance explained by the factors (say 80%), or by applying the scree test (Cattell, 1966), which plots the successive eigenvalues of the sample covariance (or correlation) matrix and looks for a spot where the eigenvalues abruptly level off. These two methods are usually criticized for being subjective. An alternative method, called the Guttman-Kaiser (GK) rule (Guttman, 1954; Kaiser, 1960), sets the number of factors to be the number of eigenvalues exceeding 1 based on the sample correlation matrix. Some likelihood based criteria include Akaike's information criterion (AIC) (Akaike, 1974, 1987), the Bayesian information criterion (BIC) (Schwarz, 1978), and Bozdogan's index of informational complexity criterion (ICOMP) (Bozdogan and Ramirez, 1987; Bozdogan and Shigemasu, 1998):

$$\mathrm{AIC}(q) = -2l\big(\hat{\boldsymbol{A}}_q^{(ML)}, \hat{\boldsymbol{\Gamma}}_q^{(ML)}\big) + 2m,$$

$$\mathrm{BIC}(q) = -2l\big(\hat{\boldsymbol{A}}_q^{(ML)}, \hat{\boldsymbol{\Gamma}}_q^{(ML)}\big) + m\log(n),$$

and

$$\mathrm{ICOMP}(q) = -2l\big(\hat{\boldsymbol{A}}_q^{(ML)}, \hat{\boldsymbol{\Gamma}}_q^{(ML)}\big) + 2(q+1)\left(\frac{p}{2}\log\left(\frac{\mathrm{tr}(\hat{\boldsymbol{\Gamma}}_q^{(ML)})}{p}\right) - \frac{1}{2}\log|\hat{\boldsymbol{\Gamma}}_q^{(ML)}|\right),$$

where $m \equiv p(q+1) - q(q-1)/2 + p$ is the number of free parameters. These criteria select the number of factors by minimizing their corresponding criterion values. In addition, the number of factors can also be selected by successively applying the likelihood ratio (LR) test (Lawley and Maxwell, 1971):

$$\mathrm{LR}_a(q) = \big(n-1-(2p+4q+5)/6\big)\log\left(\frac{\hat{\boldsymbol{A}}_q^{(ML)}\hat{\boldsymbol{A}}_q^{(ML)'}+\hat{\boldsymbol{\Gamma}}_q^{(ML)}}{|\boldsymbol{S}|}\right),$$

with a given significance level a. Note that $\mathrm{LR}_a(q)$ is asymptotically chi-squared distributed with $(1/2)\big((p-q)^2-(p+q)\big)$ degrees of freedom. Recently, Tu et al. (2009) proposed a new visualization method for selecting the number of components in principal component analysis, which may also be applied to select the number of factors.

However, all these methods perform well in some situations but poorly in others, and hence are not completely satisfactory. For example, the GK rule tends to select too many factors, particularly for large p (Fachel, 1986). The LR test depends on the choice of the significance level a, and tends to select more factors for larger sample sizes (Dhrymes et al., 1984). Both AIC and BIC are designed only for maximum likelihood factor analysis. Clearly, AIC tends to select more factors than BIC by having a smaller penalty. In addition, ICOMP suffers from a tendency to select too many factors (Lopes and West, 2004).

In this article, an approach based on unbiased risk estimation is introduced to select the number of factors, which utilizes a concept called generalized degrees of freedom (GDF) (Ye, 1998; Huang and Chen, 2007) originally proposed for model selection in regression and geostatistical regression. A data perturbation technique is applied to estimate GDF. This criterion is applicable to arbitrary estimates $\hat{\boldsymbol{\alpha}}_q$, $\hat{\boldsymbol{A}}_q$, $\hat{\boldsymbol{\Gamma}}_q$ and $\{\hat{\boldsymbol{f}}_{q,i}\}$ regardless of whether they are obtained by the principal factor method, the maximum likelihood method, or even an iterative procedure. Additionally, simulation results show that our method generally performs better than existing methods in selecting the true number of factors.

The rest of this paper is organized as follows. In Section 2, we define GDF for the orthogonal factor model and introduce our method. Section 3 provides some numerical experiments. Section 4 applies this method to a real data set, which consists of 26 intelligence tests. Finally, a brief discussion is given in Section 5.

2. The method

Consider the orthogonal factor model of (1). Let $\boldsymbol{X} \equiv (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$, and for $q = 1, \ldots, q_{\max}$, let $\hat{\boldsymbol{X}}_q \equiv \big(\hat{\boldsymbol{X}}_{q,1}, \ldots, \hat{\boldsymbol{X}}_{q,n}\big)'$ be an estimate of $\boldsymbol{X}$ based on q factors. Applying the following identity (Efron, 2004) for $q = 1, \ldots, q_{\max}$ and $i = 1, \ldots, n$:

$$\big(\hat{\boldsymbol{X}}_{q,i}-\boldsymbol{X}_i\big)'\big(\hat{\boldsymbol{X}}_{q,i}-\boldsymbol{X}_i\big) = \big(\boldsymbol{Y}_i-\hat{\boldsymbol{X}}_{q,i}\big)'\big(\boldsymbol{Y}_i-\hat{\boldsymbol{X}}_{q,i}\big) + 2\big(\hat{\boldsymbol{X}}_{q,i}-\boldsymbol{X}_i\big)'(\boldsymbol{Y}_i-\boldsymbol{X}_i) - (\boldsymbol{Y}_i-\boldsymbol{X}_i)'(\boldsymbol{Y}_i-\boldsymbol{X}_i),$$

the averaged MSE, corresponding to the risk of (3), can be written as

$$Q(q)=E(L(q))=\frac{1}{n}\left\{\sum_{i=1}^{n}E\big(\boldsymbol{Y}_i-\hat{\boldsymbol{X}}_{q,i}\big)'\big(\boldsymbol{Y}_i-\hat{\boldsymbol{X}}_{q,i}\big)\right\}+\frac{2}{n}\left\{\sum_{j=1}^{p}\gamma_jD_j(q)\right\}-\mathrm{tr}(\boldsymbol{\Gamma}),$$

where for $j=1,\ldots,p$,

$$D_j(q)\equiv\frac{1}{\gamma_j}\sum_{i=1}^{n}E\big((\hat{x}_{q,i,j}-x_{q,i,j})(y_{i,j}-x_{i,j})\big)=\frac{1}{\gamma_j}\sum_{i=1}^{n}E\big(\hat{x}_{q,i,j}(y_{i,j}-x_{i,j})\big)=\frac{1}{\gamma_j}\sum_{i=1}^{n}E\big(cov(\hat{x}_{q,i,j},y_{i,j}|\boldsymbol{X})\big),$$

$x_{i,j}, y_{i,j}$ and $\hat{x}_{q,i,j}$ are the jth elements of $\boldsymbol{X}_i$, $\boldsymbol{Y}_i$ and $\hat{\boldsymbol{X}}_{q,i}$, and γ_j is the jth diagonal element of $\boldsymbol{\Gamma}$. Suppose that $\sum_{i=1}^{n}E(|\hat{x}_{q,i,j}||\boldsymbol{X})<\infty$ almost surely. It follows from Lemma 1 of Huang and Chen (2007) that $\frac{\partial}{\partial x_{i,j}}E\big(\hat{x}_{q,i,j}|\boldsymbol{X}\big)=\frac{1}{\gamma_j}cov(\hat{x}_{q,i,j},y_{i,j}|\boldsymbol{X})$. Therefore, $D_j(q)$ can also be expressed as

$$D_j(q)=\sum_{i=1}^{n}E\left(\frac{\partial}{\partial x_{i,j}}E\big(\hat{x}_{q,i,j}|\boldsymbol{X}\big)\right). \tag{4}$$

Following Ye (1998) and Huang and Chen (2007), we call $D_j(q)$ in (4) the GDF for the jth component (column) of $\hat{\boldsymbol{X}}_q$, where $j=1,\ldots,p$. From (4), $D_j(q)$ can also be interpreted as the sensitivity of the jth fitted component based on q factors. Clearly, $D_j(q)$ tends to be larger for larger q due to higher fitting ability. Compared with the GDF defined in Ye (1998), where $\{x_{i,j}\}$ are fixed, this $D_j(q)$ further averages out random fluctuation of $\{x_{i,j}\}$ by taking the outer expectation in (4).

A data perturbation method of Ye (1998) is applied to estimate $D_j(q)$'s. First, the perturbed pseudo variables are generated according to

$$\boldsymbol{Y}_i^*\equiv\boldsymbol{Y}_i+\tau\boldsymbol{\delta}_i;\quad i=1,\ldots,n, \tag{5}$$

where $\tau>0$ is the perturbation size, and $\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_n\sim N(\boldsymbol{0},\boldsymbol{\Gamma})$ are independent and are independent of $\boldsymbol{Y}$. Then the uniqueness matrix of $\boldsymbol{Y}_i^*$ becomes $var(\boldsymbol{Y}_i^*|\boldsymbol{X}_i)=(1+\tau^2)\boldsymbol{\Gamma}$. For $q=1,\ldots,q_{\max}$ and $i=1,\ldots,n$, let $\boldsymbol{X}_{q,i}^*\equiv\hat{\boldsymbol{\alpha}}^*+\hat{\boldsymbol{A}}_q^*\hat{\boldsymbol{f}}_{q,i}^*$, where $\hat{\boldsymbol{\alpha}}^*$, $\hat{\boldsymbol{A}}_q^*$ and $\{\hat{\boldsymbol{f}}_q^*\}$ are estimates of $\boldsymbol{\alpha}$, $\boldsymbol{A}_q$ and $\{\boldsymbol{f}_i\}$ based on perturbed data $\boldsymbol{Y}^*=(\boldsymbol{Y}_1^*,\ldots,\boldsymbol{Y}_n^*)'$ and the q-factor model with the uniqueness matrix $(1+\tau^2)\boldsymbol{\Gamma}$. Motivated from (4), our estimate of $D_j(q)$ is given by

$$\hat{D}_j(q)=\sum_{i=1}^{n}\frac{\partial}{\partial y_{i,j}}E\big(\hat{x}_{q,i,j}^*|\boldsymbol{Y}\big)=\sum_{i=1}^{n}\frac{1}{\tau^2\gamma_j}cov\big(\hat{x}_{q,i,j}^*,y_{i,j}^*|\boldsymbol{Y}\big);\quad q=1,\ldots,q_{\max},j=1,\ldots,p, \tag{6}$$

where $\hat{x}_{q,i,j}^*$ is the jth element of $\hat{\boldsymbol{X}}_{q,i}^*$, $y_{i,j}^*$ is the jth element of $\boldsymbol{Y}_i^*$, and the second equality of (6) follows from Lemma 1 of Huang and Chen (2007). Using the same arguments as in Theorem 1 of Huang and Chen (2007), we have $E\big(\hat{D}_j(q)\big)\to D_j(q)$ as $\tau\to 0$, for $q=1,\ldots,q_{\max}$ and $j=1,\ldots,p$. Although $\hat{D}_j(q)$ is unbiased as $\tau\to 0$, it tends to have a larger variance for smaller τ. Based on our simulation experience, we recommend setting $\tau\approx 0.5$ to trade off some bias for smaller variance. A simulation experiment in Section 3 shows that the model selection results are not sensitive to the choice of τ.

In general, the GDF estimate in (6) can be computed using a Monte Carlo (MC) method by sampling perturbed data $\boldsymbol{Y}^{*(1)},\ldots,\boldsymbol{Y}^{*(M)}$ with a sufficiently large MC sample size M. That is, $\hat{D}_j(q)$ can be approximated by

$$\frac{1}{\tau^2\gamma_j(M-1)}\sum_{i=1}^{n}\sum_{m=1}^{M}\hat{x}_{q,i,j}^{*(m)}\big(y_{i,j}^{*(m)}-\bar{y}_{i,j}^*\big);\quad q=1,\ldots,q_{\max}, \tag{7}$$

where $\hat{x}_{q,i,j}^{*(m)}$ is the estimate of $x_{i,j}$ based on perturbed variables $\boldsymbol{Y}^{*(m)}$ and the q-factor model with the uniqueness matrix $(1+\tau^2)\boldsymbol{\Gamma}$, and $\bar{y}_{i,j}^*\equiv\frac{1}{M}\sum_{m=1}^{M}y_{i,j}^{*(m)}$. Therefore, an approximately unbiased estimate of the risk $Q(q)$ is

$$\tilde{Q}(q)=\frac{1}{n}\left\{\sum_{i=1}^{n}\big(\boldsymbol{Y}_i-\hat{\boldsymbol{X}}_{q,i}\big)'\big(\boldsymbol{Y}_i-\hat{\boldsymbol{X}}_{q,i}\big)\right\}+\frac{2}{n}\left\{\sum_{j=1}^{p}\gamma_j\hat{D}_j(q)\right\}-\mathrm{tr}(\boldsymbol{\Gamma});\quad q=1,\ldots,q_{\max}, \tag{8}$$

if $\boldsymbol{\Gamma}$ is known. When $\boldsymbol{\Gamma}$ is unknown, it has to be replaced by some estimate, independent of $q\in\{1,\ldots,q_{\max}\}$, and perturbed data have to be generated based on this estimate. We suggest using the following estimate (Guttman, 1956):

$$\hat{\boldsymbol{\Gamma}}\equiv\big(diag\big(\boldsymbol{S}^{-1}\big)\big)^{-1}, \tag{9}$$

where $diag\big(\boldsymbol{S}^{-1}\big)$ is a diagonal matrix that has the same diagonal elements as $\boldsymbol{S}^{-1}$. As shown in Guttman (1956) and Krijnen (2006), $\hat{\boldsymbol{\Gamma}}\to\boldsymbol{\Gamma}$ in probability, as $n\to\infty$ and $p\to\infty$. Therefore, the risk $Q(q)$ can be estimated by

$$\hat{Q}(q)=\frac{1}{n}\left\{\sum_{i=1}^{n}\big(\boldsymbol{Y}_i-\hat{\boldsymbol{X}}_{q,i}\big)'\big(\boldsymbol{Y}_i-\hat{\boldsymbol{X}}_{q,i}\big)\right\}+\frac{2}{n}\left\{\sum_{j=1}^{p}\hat{\gamma}_j\hat{D}_j(q)\right\}-\mathrm{tr}(\hat{\boldsymbol{\Gamma}});\quad q=1,\ldots,q_{\max}. \tag{10}$$

Finally, the number of factors selected by our GDF method is $\hat{q}_{\mathrm{GDF}}\equiv\arg\min_{1\le q\le q_{\max}}\hat{Q}(q)$ when $\boldsymbol{\Gamma}$ is estimated by $\hat{\boldsymbol{\Gamma}}$ in (9).

3. Simulations

In this section, we compare the performance of our method with some selection methods described in Section 1 in terms of both selection accuracy and the averaged MSE:

$$\frac{1}{n}\sum_{i=1}^{n} E\big(\hat{\boldsymbol{X}}_{\hat{q},i} - \boldsymbol{X}_i\big)'\big(\hat{\boldsymbol{X}}_{\hat{q},i} - \boldsymbol{X}_i\big), \tag{11}$$

where $\hat{q}$ is a statistic taking values in $\{1, 2, \dots, q_{\max}\}$ corresponding to the selected number of factors from a model selection method. Note that the averaged MSE is usually minimized when $\hat{q}$ equals the true number of factors, unless the true model includes comparatively weak factors.

3.1. Setup

We consider two factor loading models parameterized by $a_1, \dots, a_3$ and $b_1, \dots, b_6$:

$$\boldsymbol{M}_1(a_1, a_2, a_3) \equiv \frac{1}{\sqrt{p}}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix} diag(a_1, \dots, a_3) \otimes \boldsymbol{1}_{p/4},$$

and

$$\boldsymbol{M}_2(b_1, b_2, \dots, b_6) \equiv \frac{1}{\sqrt{p}}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 \end{pmatrix} diag(b_1, \dots, b_6) \otimes \boldsymbol{1}_{p/8},$$

where $\otimes$ is the Kronecker product, and $\boldsymbol{1}_m$ is an $m \times 1$ vector of 1. From these two models, we consider the following four cases for factor loading matrices:

Case I. $\boldsymbol{A}_{1,k} \equiv \sqrt{k}\boldsymbol{M}_1\big(\sqrt{3}, \sqrt{2}, 1\big)$;
Case II. $\boldsymbol{A}_{2,k} \equiv \sqrt{k}\boldsymbol{M}_1\big(\sqrt{2}, \sqrt{2}, \sqrt{2}\big)$;
Case III. $\boldsymbol{A}_{3,k} \equiv \sqrt{k}\boldsymbol{M}_2\big(\sqrt{6}, \sqrt{5}, \sqrt{4}, \sqrt{3}, \sqrt{2}, 1\big)$;
Case IV. $\boldsymbol{A}_{4,k} \equiv \sqrt{k}\boldsymbol{M}_2\big(\sqrt{2}, \dots, \sqrt{2}\big)$,

where $k > 0$ is used to control the signal to noise ratio, SNR $= \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}')/\mathrm{tr}(\boldsymbol{\Gamma})$. For each case, we consider $(p, n) \in \{(16, 50), (16, 200), (16, 500), (24, 200), (24, 500), (32, 200), (32, 500)\}$ and three SNRs ($k = 1, 2, 3$), where the SNRs corresponding to Cases I–IV were $3k/8$, $3k/8$, $21k/16$ and $12k/16$. The data $\boldsymbol{Y}$ were generated according to (1) with uniqueness matrix $\boldsymbol{\Gamma} = \boldsymbol{I}_p$ and mean $\boldsymbol{\alpha} = \boldsymbol{0}$. Note that the true number of factors is $q_{true} = 3$ in Cases I–II, and $q_{true} = 6$ in Cases III–IV.

3.2. Comparison among various methods

Our goal is to select an appropriate q among $\{1, 2, \dots, q_{\max}\}$ with $q_{\max} = 8$. For each q, we consider the estimate $\hat{\boldsymbol{X}}_{q,i} = \bar{\boldsymbol{Y}} + \hat{\boldsymbol{A}}_q\hat{\boldsymbol{f}}_{q,i}$ of $\boldsymbol{X}_i$; $i = 1, \dots, n$ based on ML, where $\hat{\boldsymbol{A}}_q$ is the ML estimate obtained from the Varimax factor rotation criterion (Kaiser, 1958), and $\hat{\boldsymbol{f}}_{q,i}$'s are obtained from (2). The performance of various selection methods is compared using the averaged MSE criterion of (11) under different scenarios based on 5000 simulation replicates. We compare the performance of the GDF method with various selection methods, including AIC, BIC, ICOMP, two LR tests with significance levels $a = 0.01$ and 0.05, and the GK rule.

The R function "factanal" was applied to compute the ML estimate $\hat{\boldsymbol{A}}_{\hat{q}}$. This function applies the sample correlation matrix $\boldsymbol{R} \equiv diag(\boldsymbol{S})^{-1/2}\boldsymbol{S}\, diag(\boldsymbol{S})^{-1/2}$ rather than $\boldsymbol{S}$, for computing the ML estimates, and allows users to specify the starting values for the uniqueness matrix corresponding to $\boldsymbol{R}$. We chose $(1 - 0.5\, q/p)\big(diag\big(\boldsymbol{R}^{-1}\big)\big)^{-1}$ as the starting values as suggested by Jöreskog (1963). However, these starting values may not always achieve convergence, particularly when q is close to $q_{\max} = 8$, at which the ML estimates become unstable. When convergence fails, the starting values of the uniqueness matrix are successively replaced by $\theta\big(diag\big(\boldsymbol{R}^{-1}\big)\big)^{-1}$ for a randomly generated $\theta \sim Unif(0, 2)$ until convergence is achieved. Convergence was achieved eventually in all the simulation. Clearly, some ML estimates obtained from this procedure may only be local maxima and not global maxima, resulting in an approximate ML estimation procedure. In fact, the GDF estimate $\hat{\boldsymbol{D}}_j(q)$ of (7) actually measured the sensitivity of this approximate ML estimation procedure, because the same ML estimation

Table 1
Averaged MSEs and percentages (in parentheses) of selecting the true number of factors for various model selection methods under $p = 16$ based on 5000 simulation replicates.

Case	Method	$n = 50$			$n = 500$		
		$k = 1$	$k = 2$	$k = 3$	$k = 1$	$k = 2$	$k = 3$
I	True	4.353	4.104	4.051	2.087	2.473	2.648
	Ideal	3.737 (20.72)	4.033 (85.54)	4.045 (98.58)	2.087 (100.00)	2.473 (100.00)	2.648 (100.00)
	GDF	4.155 (41.44)	4.207 (82.70)	4.147 (91.86)	2.087 (100.00)	2.473 (100.00)	2.648 (100.00)
	AIC	4.245 (41.88)	4.512 (68.68)	4.532 (70.30)	2.190 (79.58)	2.574 (79.30)	2.747 (79.02)
	BIC	4.297 (0.32)	4.551 (29.62)	4.250 (80.74)	2.090 (99.12)	2.473 (100.00)	2.648 (100.00)
	ICOMP	5.438 (19.78)	5.640 (9.64)	5.828 (5.08)	2.779 (2.60)	3.179 (1.70)	3.383 (1.16)
	$LR_{0.01}$	4.178 (3.46)	4.536 (30.42)	4.425 (71.02)	2.092 (99.00)	2.480 (98.84)	2.655 (98.84)
	$LR_{0.05}$	4.092 (11.54)	4.381 (51.90)	4.270 (82.62)	2.114 (95.04)	2.501 (94.72)	2.674 (94.68)
	GK	6.702 (0.00)	5.283 (14.96)	4.359 (73.04)	2.092 (98.02)	2.473 (100.00)	2.648 (100.00)
II	True	4.289	4.029	4.036	2.165	2.545	2.709
	Ideal	4.018 (57.56)	4.025 (98.84)	4.036 (99.96)	2.165 (100.00)	2.545 (100.00)	2.709 (100.00)
	GDF	4.379 (63.24)	4.136 (90.82)	4.121 (93.76)	2.165 (100.00)	2.545 (100.00)	2.709 (100.00)
	AIC	4.493 (57.40)	4.496 (70.98)	4.534 (70.44)	2.267 (79.44)	2.645 (79.18)	2.807 (78.98)
	BIC	4.903 (1.54)	4.395 (74.46)	4.061 (98.60)	2.165 (100.00)	2.545 (100.00)	2.709 (100.00)
	ICOMP	5.462 (17.64)	5.683 (7.86)	5.844 (4.18)	2.856 (2.20)	3.259 (1.68)	3.440 (1.28)
	$LR_{0.01}$	4.632 (7.06)	4.556 (64.68)	4.163 (93.98)	2.169 (99.00)	2.550 (98.90)	2.715 (98.86)
	$LR_{0.05}$	4.494 (20.76)	4.307 (79.66)	4.147 (93.50)	2.189 (95.02)	2.570 (94.66)	2.735 (94.56)
	GK	6.762 (0.00)	5.298 (13.00)	4.359 (72.46)	2.169 (98.04)	2.545 (100.00)	2.709 (100.00)
III	True	8.162	8.038	7.952	4.690	5.290	5.546
	Ideal	7.575 (25.78)	7.917 (75.42)	7.932 (94.22)	4.690 (99.94)	5.290 (99.98)	5.546 (99.98)
	GDF	7.752 (27.26)	8.013 (71.86)	7.971 (92.90)	4.702 (96.42)	5.290 (100.00)	5.546 (100.00)
	AIC	8.089 (42.62)	8.372 (67.74)	8.388 (66.54)	4.825 (78.46)	5.412 (77.86)	5.662 (77.92)
	BIC	9.772 (0.22)	8.517 (35.76)	8.123 (83.82)	4.690 (99.84)	5.290 (100.00)	5.546 (100.00)
	ICOMP	7.900 (39.64)	8.162 (67.00)	8.171 (68.58)	4.921 (36.92)	5.550 (29.84)	5.836 (25.00)
	$LR_{0.01}$	8.415 (1.96)	8.576 (27.60)	8.384 (69.70)	4.697 (99.00)	5.296 (98.92)	5.552 (98.86)
	$LR_{0.05}$	8.133 (7.76)	8.335 (50.00)	8.145 (84.32)	4.723 (95.26)	5.320 (94.78)	5.575 (94.68)
	GK	7.781 (18.54)	8.787 (7.34)	10.077 (5.12)	5.162 (0.72)	6.778 (0.02)	8.487 (0.00)
IV	True	8.305	8.041	7.913	4.303	5.057	5.382
	Ideal	7.753 (30.22)	7.980 (85.82)	7.902 (96.42)	4.303 (99.92)	5.057 (99.98)	5.382 (99.98)
	GDF	8.050 (33.92)	8.055 (84.80)	7.924 (96.20)	4.303 (100.00)	5.057 (100.00)	5.382 (100.00)
	AIC	8.341 (28.92)	8.372 (68.26)	8.360 (66.48)	4.447 (78.22)	5.183 (77.90)	5.502 (77.82)
	BIC	10.654 (0.00)	11.276 (23.26)	8.243 (85.96)	4.303 (100.00)	5.057 (100.00)	5.382 (100.00)
	ICOMP	8.008 (35.74)	8.121 (69.44)	8.064 (74.88)	4.515 (40.20)	5.289 (33.92)	5.627 (31.12)
	$LR_{0.01}$	8.958 (0.56)	9.296 (22.08)	8.614 (69.62)	4.311 (99.06)	5.063 (99.00)	5.388 (98.94)
	$LR_{0.05}$	8.630 (3.64)	8.730 (46.06)	8.194 (85.32)	4.338 (95.26)	5.088 (94.88)	5.412 (94.64)
	GK	8.307 (81.62)	8.053 (93.80)	7.960 (96.38)	4.303 (100.00)	5.057 (100.00)	5.382 (100.00)

procedure is applied to perturbed variables $\boldsymbol{Y}^*$ in computing $\hat{\boldsymbol{D}}_j(q)$, with the perturbation size $\tau = 0.5$ and the MC sample size $M = 100$ in (7).

3.3. Simulation results

The averaged MSE values of various selection methods for the four cases (factor loading matrices) are displayed in Table 1 for various SNRs under $n = 50, 500$, and $p = 16$, and in Table 2 for various p values under $n = 200$ and $k = 1$. The results based on the true number of factors (denoted as "True") and the ideal number of factors, $\hat{q}^* \equiv \arg\min_{1 \le q \le q_{\max}} L(q)$ (denoted as "Ideal"), for each simulation replicate are also shown. Note that the true number of factors does not always correspond to the smallest loss, particularly when SNR is small. The proportions of times for selecting the true number of factors are also reported in Tables 1 and 2. Moreover, the proportions of models that were underfitted and overfitted by one factor are also reported in Table 2. It can be seen that the GDF method performs the best or close to the best in almost all cases in terms of the averaged MSE. Our GDF method also performs the best or close to the best in most of the cases in terms of selecting the true number of factors.

Although BIC performs well for large sample sizes (e.g., $n = 500$), it performs poorly for small ones (e.g., $n = 50$), particularly when SNRs are small. The reason is that BIC tends to underestimate the true number of factors when the sample size is not large enough. On the other hand, AIC tends to overestimate the true number of factors. Comparing the averaged MSE values between AIC and BIC, the averaged MSE appears to be more affected by underestimation than overestimation. As expected, ICOMP tends to overestimate the true number of factors, and it has an even more serious overestimation problem than AIC, resulting in large averaged MSE values for Cases I and II, where $q_{true} = 3$. Similar to BIC, the LR methods appear to perform well for large sample sizes, but poorly for small ones, supporting the suggestion that the LR method should only be used when $n - p \ge 50$ (Bartlett, 1950). Comparing the two LR methods with different significance levels, $LR_{0.05}$ performs

Table 2
Percentages of selected numbers of factors and the averaged MSEs for various model selection methods under $k = 1$ and $n = 200$ based on 5000 simulation replicates, where the numbers given in parentheses are the corresponding standard errors.

p	Method	$q_{true} - 1$	q_{true}	$q_{true} + 1$	MSE	$q_{true} - 1$	q_{true}	$q_{true} + 1$	MSE
		Case I				Case II			
16	GDF	7.64	92.32	0.04	2.390 (0.033)	0.00	99.98	0.02	2.429 (0.029)
	AIC	0.34	79.60	17.90	2.527 (0.080)	0.00	78.78	18.76	2.589 (0.080)
	BIC	69.98	29.66	0.00	2.612 (0.046)	2.78	97.18	0.00	2.452 (0.040)
	$LR_{0.01}$	26.34	73.06	0.52	2.461 (0.044)	0.36	98.86	0.70	2.438 (0.035)
	$LR_{0.05}$	11.26	84.90	3.50	2.436 (0.050)	0.02	95.60	3.90	2.465 (0.051)
24	GDF	8.52	90.60	0.88	2.614 (0.036)	0.00	99.52	0.48	2.641 (0.033)
	AIC	1.58	83.70	13.28	2.708 (0.070)	0.00	84.02	14.18	2.753 (0.071)
	BIC	91.20	2.42	0.00	2.923 (0.058)	24.92	72.12	0.00	2.901 (0.098)
	$LR_{0.01}$	56.60	42.62	0.68	2.750 (0.046)	4.68	94.50	0.74	2.683 (0.051)
	$LR_{0.05}$	32.04	63.68	3.76	2.709 (0.054)	0.80	94.46	4.14	2.680 (0.052)
32	GDF	7.82	88.68	3.38	2.858 (0.042)	0.02	97.84	2.06	2.871 (0.038)
	AIC	3.66	85.90	9.86	2.910 (0.060)	0.00	88.26	11.00	2.938 (0.060)
	BIC	69.54	0.08	0.00	3.304 (0.095)	47.14	26.22	0.00	3.745 (0.139)
	$LR_{0.01}$	72.82	26.06	0.78	2.971 (0.042)	18.20	80.82	0.86	3.011 (0.073)
	$LR_{0.05}$	49.36	46.42	3.34	2.949 (0.054)	6.46	88.94	3.66	2.944 (0.064)
		Case III				Case IV			
16	GDF	50.60	49.30	0.00	5.302 (0.059)	1.72	98.24	0.00	4.843 (0.059)
	AIC	0.12	79.28	18.84	5.389 (0.110)	0.00	78.86	19.36	5.065 (0.119)
	BIC	52.66	47.22	0.00	5.339 (0.061)	14.1	80.16	0.00	5.080 (0.134)
	$LR_{0.01}$	15.74	83.70	0.54	5.234 (0.061)	2.66	96.88	0.46	4.861 (0.066)
	$LR_{0.05}$	5.90	90.38	3.50	5.242 (0.076)	0.42	95.80	3.64	4.884 (0.079)
24	GDF	23.80	76.18	0.02	5.487 (0.051)	0.46	99.46	0.08	5.108 (0.048)
	AIC	0.84	80.38	16.74	5.591 (0.085)	0.00	80.08	17.96	5.261 (0.088)
	BIC	89.82	6.06	0.00	5.746 (0.064)	30.62	25.56	0.00	6.449 (0.258)
	$LR_{0.01}$	48.86	50.42	0.70	5.574 (0.056)	19.08	80.16	0.72	5.255 (0.077)
	$LR_{0.05}$	26.48	69.88	3.16	5.539 (0.062)	6.20	89.94	3.42	5.180 (0.068)
32	GDF	18.92	80.66	0.42	5.817 (0.051)	0.26	99.24	0.50	5.476 (0.050)
	AIC	3.24	82.64	12.78	5.905 (0.077)	0.02	85.72	12.82	5.581 (0.078)
	BIC	74.70	0.16	0.00	6.234 (0.096)	5.50	1.12	0.00	8.722 (0.272)
	$LR_{0.01}$	70.24	28.82	0.62	5.942 (0.055)	42.64	56.20	0.64	5.783 (0.087)
	$LR_{0.05}$	45.28	50.82	3.16	5.914 (0.063)	20.62	75.52	3.18	5.643 (0.080)

better when $n = 50$, whereas $LR_{0.01}$ performs better when $n = 200$ and 500. Although the GK rule performs well in Case IV, it performs miserably in Case III.

3.4. The perturbation size τ

We performed a simulation experiment regarding the sensitivity of the perturbation size τ. We consider Case III with $k = 2$ fitted by $q = 4$ factors. Fig. 1 shows the boxplots of the estimated penalty $\sum_{j=1}^{p} \gamma_j \hat{D}_j(q)$ in (8) with respect to $\tau \in \{0.01, 0.1, 0.3, 0.5, 0.7, 0.9\}$ for various sample sizes. As we expect, a larger τ corresponds to a larger bias but a smaller variance, and vice versa. Similar results can be seen for other cases. Although $\sum_{j=1}^{p} \gamma_j \hat{D}_j(q)$ can be seen to be almost unbiased when τ is 0.01, it tends to have a very large variance. Fig. 2 shows the boxplots of the corresponding penalty estimation errors, $\sum_{j=1}^{p} \gamma_j (\hat{D}_j(q) - D_j)$, against $q = 1, \ldots, q_{\max}$ at $\tau = 0.01, 0.3$ and 0.5. Notice that the biases tend to be a constant when q is around the true number of factors (i.e., $q_{true} = 6$) for each case. As the biases corresponding to $q = 4, 5, 6, 7, 8$ are similar, selection based on minimizing $\hat{Q}(q)$ in (8) is not much affected by the bias. The results justify the choice of $\tau = 0.5$ over $\tau = 0.01$ based on the fact that bias in $\hat{D}_j(q)$ essentially has no effect on model selection, but the variance of $\hat{D}_j(q)$ for $\tau = 0.5$ is much smaller than that for $\tau = 0.01$. Similar results can be seen for other cases.

4. Application

We applied the orthogonal factor model of (1) to an intelligence data set from a study by Holzinger and Swineford (1939). This study contained 26 tests designed with five phases of measurements: spatial, verbal, mental speed, memory, and mathematical ability. The tests were taken by 301 students in the 7th and 8th grades from two Chicago schools. We considered only 24 tests, because the other two tests contained 156 missing values and were known to be similar to two of the 24 tests. Here, we consider two factor rotation criteria: orthogonal Varimax (Kaiser, 1958) and oblique Promax (Hendrickson and White, 1964), along with two factor score estimation methods: the generalized least squares method (Bartlett, 1937, 1938) and the one given in (2), resulting in a combination of four methods, which we denote by $\kappa = 1, \ldots, 4$.

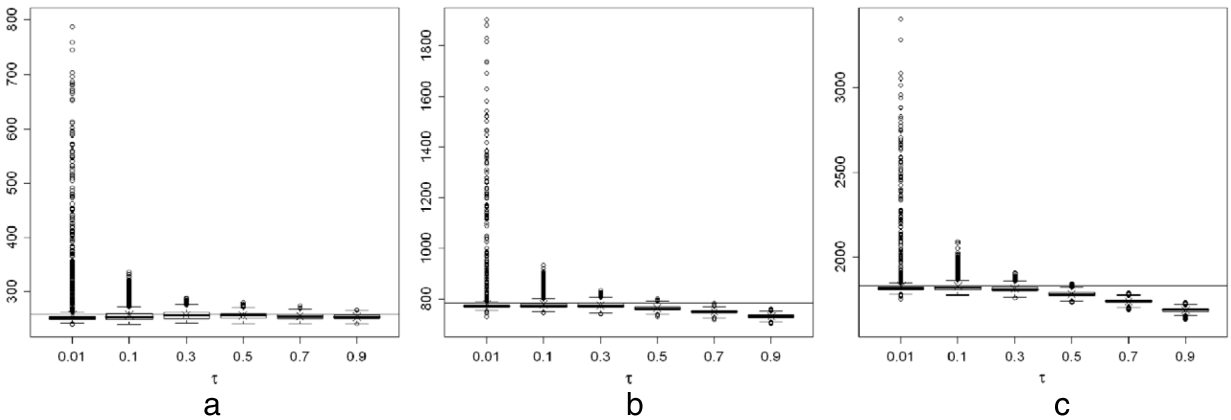

Fig. 1. Boxplots of the estimated penalty, $\sum_{j=1}^{p} \gamma_j \hat{D}_j(q)$, for Case III with $k = 2$ and $q = 4$ with various perturbation sizes τ based on 5000 simulation replicates: (a) $n = 50$; (b) $n = 200$; (c) $n = 500$.

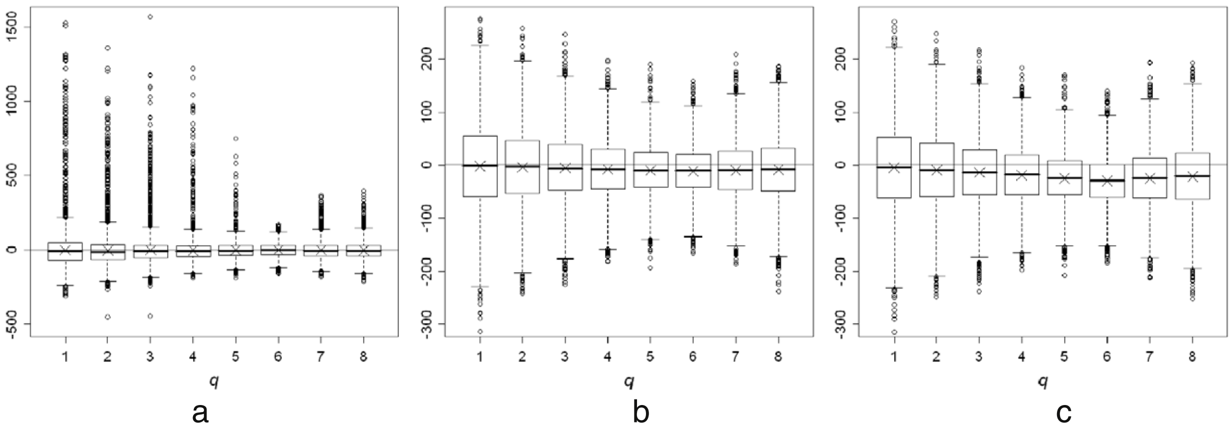

Fig. 2. Boxplots of the penalty estimation errors, $\sum_{j=1}^{p} \gamma_j (\hat{D}_j(q) - D_j)$, for Case III with $k = 2$ and $n = 200$ with respect to various numbers of factors q based on 5000 simulation replicates: (a) $\tau = 0.01$; (b) $\tau = 0.3$; (c) $\tau = 0.5$.

As in the simulation experiment, we consider $q_{\max} = 8$. Denote the estimate of $\boldsymbol{X}_i$ fitted by method κ and the q-factor model as $\hat{\boldsymbol{X}}_{\kappa,q,i} = \bar{\boldsymbol{Y}} + \hat{\boldsymbol{A}}_{\kappa,q} \hat{\boldsymbol{f}}_{\kappa,q,i}$, for $\kappa = 1, \ldots, 4$, $q = 1, \ldots, q_{\max}$ and $i = 1, \ldots, n$. Our risk estimate is given by

$$\hat{Q}(\kappa, q) = \frac{1}{n} \left\{ \sum_{i=1}^{n} \left(\boldsymbol{Y}_i - \hat{\boldsymbol{X}}_{\kappa,q,i} \right)' \left(\boldsymbol{Y}_i - \hat{\boldsymbol{X}}_{\kappa,q,i} \right) \right\} + \frac{2}{n} \left\{ \sum_{j=1}^{p} \hat{\gamma}_j \hat{D}_j(\kappa, q) \right\} - \operatorname{tr}(\hat{\boldsymbol{\Gamma}}); \quad q = 1, \ldots, q_{\max},$$

where

$$\hat{D}_j(\kappa, q) \equiv \frac{1}{\tau^2 \hat{\gamma}_j (M-1)} \sum_{i=1}^{n} \sum_{m=1}^{M} \hat{x}^{*(m)}_{\kappa,q,i,j} \left(y^{*(m)}_{i,j} - \bar{y}^{*}_{i,j} \right); \quad q = 1, \ldots, q_{\max},$$

$\tau = 0.5$ and $M = 100$.

The smallest value of $\hat{Q}(\kappa, q)$ corresponds to the Varimax criterion with factor scores estimated by (2) based on the four-factor model. The resulting factor loadings based on our selected number of factors are displayed in Table 3. Basically, the first factor concentrating on tests T5–T9 is a verbal measure. The second factor is mainly for spatial and mathematical ability. The third factor corresponds to a mental speed measure and the fourth factor is related to memory ability. In fact, the five-factor model is equally competitive by further separating the second factor into one additional factor corresponding to mathematical ability. However, this new factor has relatively small factor loadings that are not larger than those of the other factors even for mathematical variables. Both BIC and the GK rule also select four factors, and $\text{LR}_{0.01}$ selects five factors. On the other hand, AIC and $\text{LR}_{0.05}$ select six factors. We found some of these factors more difficult to interpret.

5. Discussion

In this article, we introduce a novel selection method for selecting the number of factors, which performs better than some commonly used methods in our simulation, despite being more computationally intensive. Although we only consider

Table 3
Estimated factor loadings for Holzinger and Swineford's data set.

Phases	Tests	Factor 1	Factor 2	Factor 3	Factor 4
Spatial	T1 (visual perception)	0.239	0.614	0.141	0.143
	T2 (cubes)		0.515		
	T3 (paper form board)	0.146	0.431		
	T4 (lozenges)		0.618	0.126	0.174
Verbal	T5 (general information)	0.816	0.140	0.148	
	T6 (paragraph comprehension)	0.782	0.171	0.109	0.150
	T7 (sentence completion)	0.866	0.115		
	T8 (word classification)	0.690	0.216	0.124	0.128
	T9 (word meaning)	0.805	0.236		0.111
Mental speed	T10 (addition)		−0.105	0.764	0.150
	T11 (code)	0.256	0.129	0.553	0.263
	T12 (counting)		0.190	0.648	
	T13 (straight and curved capitals)	0.112	0.367	0.517	
Memory	T14 (word)	0.135			0.651
	T15 (number)		0.122		0.564
	T16 (figure)	0.134	0.377	0.106	0.499
	T17 (object-number)			0.298	0.539
	T18 (number-figure)		0.140	0.210	0.431
	T19 (figure-word)	0.219	0.193	0.120	0.352
Mathematical ability	T20 (deduction)	0.299	0.462		0.270
	T21 (numerical puzzle)	0.253	0.395	0.378	0.205
	T22 (problem reasoning)	0.446	0.438	0.105	0.164
	T23 (series completion)	0.368	0.546	0.207	0.216
	T24 (Woody-McCall mixed fundamentals)	0.385	0.213	0.370	0.294

maximum likelihood in the simulation examples, our selection method is also applicable to any other estimation methods for factor loadings and factor scores. In the real data example, we select among two factor rotation methods: orthogonally rotated Varimax and obliquely rotated promax, which are combined with two factor score estimation methods, and thus demonstrate how this method can be applied to select estimation methods of factor rotations and factor scores.

With promising simulation results, it is of interest to investigate some theoretical properties of the proposed method. One possible way could be to formulate the factor analysis problem as a least squares problem with the factor loading matrix estimated by principal components, and try to extend the results of Shao (1997) as both p and n go to infinity. However that is beyond the scope of this paper.

Acknowledgement

The research was supported in part by NSC 95-2118-M-001-016-MY2 and NSC 97-2118-M-001-009.

Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.csda.2009.10.002.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.
Akaike, H., 1987. Factor analysis and AIC. Psychometrika 52, 317–332.
Bartholomew, D.J., Knott, M., 1999. Latent Variable Models and Factor Analysis, 2nd ed. Arnold, London.
Bartlett, M.S., 1937. The statistical conception of mental factors. British Journal of Psychology 28, 97–104.
Bartlett, M.S., 1938. Methods of estimating mental factors. Nature 141, 609–610.
Bartlett, M.S., 1950. Tests of significance in factor analysis. British Journal of Psychology (Statistical Section) 3, 77–85.
Bozdogan, H., Ramirez, D.E., 1987. An expert model selection approach to determine the best pattern structure in factor analysis models. In: Bozdogan, H., Gupta, A.K. (Eds.), Multivariate Statistical Modeling and Data Analysis. Reidel Publishing Company, pp. 35–60.
Bozdogan, H., Shigemasu, K., 1998. Bayesian factor analysis model and choosing the number of factors using a new informational complexity criterion. Technical Report. Department of Statistics, University of Tennessee.
Cattell, R.B., 1966. The scree test for the number of factors. Multivariate Behavioral Research 1, 246–276.
Dhrymes, P.J., Friend, I., Glutekin, N.B., 1984. A critical reexamination of the empirical evidence on the arbitrage pricing theory. Journal of Finance 39, 323–346.
Efron, B., 2004. The estimation of prediction error: Covariance penalties and cross-validation (with discussion). Journal of the American Statistical Association 99, 619–642.
Fachel, G.M.G., 1986. The C-type distribution as an underlying model for categorical data and its use in factor analysis. Ph.D. thesis, University of London.
Guttman, L., 1954. Some necessary conditions for common factor analysis. Psychometrika 19, 149–162.

Guttman, L., 1956. Best possible systemic estimates of communalities. Psychometrika 21, 273–285.
Hendrickson, A.E., White, P.O., 1964. Promax: A quick method for rotation to oblique simple structure. British Journal of Statistical Psychology 17, 65–70.
Holzinger, K.J., Swineford, F., 1939. A study in factor analysis: The stability of a bi-factor solution. In: Supplementary Educational Monographs 48. University of Chicago.
Huang, H-C., Chen, C-S., 2007. Optimal geostatistical model selection. Journal of the American Statistical Association 102, 1009–1024.
Jöreskog, K.G., 1963. Statistical Estimation in Factor Analysis. Almqvist & Wilksell, Uppsala.
Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika 23, 187–200.
Kaiser, H.F., 1960. The application of electronic computer to factor analysis. Educational and Psychological Measurement 20, 141–151.
Krijnen, W.P., 2006. Convergence of estimates of unique variances in factor analysis based on the inverse sample covariance matrix. Psychometrika 71, 193–199.
Lawley, D.N., Maxwell, A.E., 1971. Factor Analysis as a Statistical Method. Butterworths, London.
Lopes, H.F., West, M., 2004. Bayesian model assessment in factor analysis. Statistica Sinica 14, 41–67.
Rencher, A.C., 2002. Methods of Multivariate Analysis. Wiley, New York.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.
Shao, J., 1997. An asymptotic theory for linear model selection. Statistica Sinica 7, 221–264.
Thomson, G.H., 1934. The meaning of i in the estimate of g. British Journal of Psychology 25, 92–99.
Tu, I.-P., Chen, H., Chen, X., 2009. An eigenvector variability plot. Statistica Sinica 19, 1741–1754.
Ye, J., 1998. On measuring and correcting the effects of data mining and model selection. Journal of the American Statistical Association 93, 120–131.