

Modified estimator of the contribution rates of population eigenvalues

Yo Sheena*

April 2010

Abstract

Modified estimators for the contribution rates of population eigenvalues are given under an elliptically contoured distribution. These estimators decrease the bias of the classical estimator, i.e. the sample contribution rates. The improvement of the modified estimators over the classical estimator is proved theoretically in view of their risks. We also checked numerically that the drawback of the classical estimator, namely the underestimation of the dimension in principal component analysis or factor analysis, are corrected in the modification.

MSC(2010) *Subject Classification*: Primary 62H12; Secondary 62H25

Key words and phrases: contribution rates, Elliptically contoured distribution, principal component analysis, factor analysis

1 Introduction

Let Σ be the population covariance matrix of a p -variate random vector $\mathbf{x} = (x_1, \dots, x_p)$. Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$, $\lambda_1 \geq \dots \geq \lambda_p$ denote the eigenvalues of Σ , then the population contribution rates are defined as

$$\boldsymbol{\tau} = (\tau_1, \dots, \tau_p), \quad \tau_i = \lambda_i / \sum_{j=1}^p \lambda_j, \quad i = 1, \dots, p. \quad (1)$$

The contribution rates play an important role in statistical linear models. Especially in principal component analysis or factor analysis, it gives an important information for the determination of the model's dimension; "How many principal components substantially represents the total variance?" is a basic quercitin in principal component analysis, and the number of factors to be incorporated in a model is a crucial problem in factor analysis. For this issue, the most simple and widely used methods are the following ones based on the population contribution rates.

*Department of Economics, Shinshu University

1. The cumulative percentage of the eigenvalues

With a cut-off t^* , we determine the smallest integer m for which

$$\sum_{i=1}^m \tau_i \geq t^* \tag{2}$$

to be the number of principal components or factors to be retained. Practically a number between 0.7 and 0.9 is often chosen as a cut-off t^* .

2. The relative size of each eigenvalue

If the i th eigenvalue is larger than the average of the population variance $\sum_{i=1}^p \lambda_i/p$, the corresponding principal component or factor is to be retained. This criteria is equivalent to check whether τ_i satisfies the inequality

$$\tau_i > p^{-1}. \tag{3}$$

This is also equivalent to “Kaiser criterion” in factor analysis, which asserts that the number of the eigenvalues larger than unit of the population *correlation* matrix should be the number of factors.

Many methods have been proposed for the determination of dimension relating to principal component analysis or factor analysis (or more generally covariance structure model). See Jolliffe (2002) and Fabrigar et al.(1999), both of which give an extensive review of the methods for choosing a dimension respectively for principal component analysis and exploratory factor analysis. According to Jolliffe (2002)’s classification, there are several categories for the methods other than that based on the population contribution rates; 1) Hypothesis testing method, 2) Information theoretic method , 3) computer-intensive method.

Furthermore it might be better to add another category, “large dimensional random matrix method”, if we could name it. For the past decade, while the results have (re)accumulated on “general asymptotics” that considers the limiting operation of both p (the dimension) and n (the sample numbers), we have seen much improvement on this method. The limiting distribution of the sample eigenvalues under general asymptotics gives some novel ideas for the dimension determination. It is appealing that several simulations show that the arguments based on general asymptotics is effective even if n and p are relatively small. See e.g. Kritchman and Nadler (2004), Ulfarsson and Solo (2008). They propose highly efficient methods for a so-called “spiked covariance model” (see the equation (38)). For large dimensional random matrix theories used in these papers, see the references therein. We also refer to Paul (2007), Nadler (2008), Karoui (2009) for more recent developments.

We should notice that the concept of “dimension” could be rather ambiguous term. In the fields such as physics or chemistry, it is often the case that there exist “signals (components)” and “noise” in its own mechanism with clear distinction. Naturally the analysis of the covariance structure is aimed at “detecting” the numbers of the signals (components), as the term “signal detection” indicates. On the contrary in psychology or economics, a “factor” in its theory is rather abstract object and sometimes impossible to draw the line between the “factors” and “noise”. There we could only say some factors are trivial while the others are nontrivial. Hence the number of the factors (i.e. dimension) are not considered to preexist before the statistical inference but rather be determined through the inference so that we can carry out dimension reduction without serious loss of information. We could say it is “deciding” the dimension. Considering the both cases, “detecting” and “deciding” the dimension, it seems that there is no single excellent method that is universally effective. After all we had better choose effective methods according to the purpose of the determination of the dimension and/or the presupposed mechanism of data generation.

Back to the methods (2), (3) of our concern, their cut-off values seem somewhat ad-hoc without rigorous theoretical background. We naturally raise a question such as “Why is 0.9 for t^* chosen ?” We only could insist that it is nothing more than conventional criteria for the inference like a given significant level (e.g. 5%) in a hypothesis test. Nevertheless, they have been widely used and incorporated into many softwares for statistical analysis because of their simplicity and easiness for calculation. They do not suppose any rigid data generation mechanism, which is often suitable for the purpose of dimension “decision (reduction)”. We think that the improved inference on the population contribution rates could make some contribution to the better dimension reduction. In this paper we focus ourselves to the point estimation of $\boldsymbol{\tau}$ using the sample covariance matrix.

Let \mathbf{A} denote the (unbiased) sample covariance matrix and $l_1^* \geq \dots \geq l_p^* > 0$ be its ordered eigenvalues. Then the sample contribution rates are defined as

$$d_i = l_i^* / \sum_{j=1}^p l_j^*, \quad 1 \leq i \leq p. \quad (4)$$

Traditionally (and perhaps almost always) the set of sample contribution rates has been used for the estimation of the population contribution rates. Hereafter the sample contribution rates as an estimator of $\boldsymbol{\tau}$ will be called “classical estimator” and denoted by $\hat{\boldsymbol{\tau}}^{(0)} = (\hat{\tau}_1^{(0)}, \dots, \hat{\tau}_p^{(0)})$, where $\hat{\tau}_i^{(0)} = d_i$, $i = 1, \dots, p$.

As far as we know, for the estimation of $\boldsymbol{\tau}$ there is no other option than the classical estimator. However $\hat{\boldsymbol{\tau}}^{(0)}$ seems to have nonnegligible bias. It is well known that the sample eigenvalues $\mathbf{l}^* = (l_1^*, \dots, l_p^*)$ are biased. Specifically

saying, λ_i , $i = 1, \dots, p$ is majorized by $E(l_i^*)$, $i = 1, \dots, p$, that is,

$$\sum_{j=1}^m \lambda_j \leq \sum_{j=1}^m E(l_j^*), \quad 1 \leq m \leq p.$$

This fact makes us conjecture that $\hat{\tau}^{(0)}$ is also biased. Let $\mathbf{S} = (N - 1)\mathbf{A}$, where N is the number of the samples. In the case \mathbf{S} is distributed as Wishart matrix and λ 's have no multiplicity, the expected value of d_i is expanded with respect to the degree of freedom $n(= N - 1)$ as follows (see the proof in Appendix);

$$E(d_i) = \tau_i + n^{-1} \left\{ 2 \frac{\lambda_i (\sum_{j=1}^p \lambda_j^2)}{(\sum_{j=1}^p \lambda_j)^3} - 2 \frac{\lambda_i^2}{(\sum_{j=1}^p \lambda_j)^2} + \frac{\lambda_i}{\sum_{j=1}^p \lambda_j} \sum_{j \neq i} \frac{\lambda_j}{\lambda_i - \lambda_j} \right\} + O(n^{-2}) \quad (5)$$

The coefficient of the n^{-1} term is complicated, but we easily notice that when λ_i 's are close to each other, large positive (negative) bias might take place for the smaller (larger) i 's. Note that the similar expansion with respect to both p and n might be possible in view of "general asymptotics". Please refer to Nadler (2008) for a matrix perturbation approach.

The simulated results of the distribution of d_i 's under the condition Σ is a identity matrix can be found in Mandel (1972) and Krzanowski (1979). They observe the large bias of d_i 's. (See also Sugiyama and Tong (1976), Konishi (1977), and Huang and Tseng (1992) for the distribution of d_i 's.) Table 1 shows the simulated values of $E(d_i)$, $i = 1, \dots, 10$ calculated from 10000 random 10-dimensional Wishart matrices with the degree of freedom 30 generated under several patterns of λ . (The total sum of λ_i 's always equals one, hence $\lambda_i = \tau_i$, $1 \leq i \leq p$.) As the figures in the table show, it is not seldom that the bias of the first or last few $E(d_i)$'s surpasses 50% (sometimes 100%) of λ_i , while the sign of the bias for the middle part of $E(d_i)$'s is quite unstable.

The aim of this paper is to derive an alternative estimator that modifies the bias of the classical estimator. In the next section, first we show the distribution of the sample contribution rates is identical under a class of elliptical distributions. Second we propose a class of new estimators and show their superiority to the classical estimator under the class of elliptical distributions from a decision theoretic point of view. In the third section, by simulation studies, we clarify other preferable aspects of the new estimator.

2 Main Result

2.1 Framework

Before deriving new estimators, we formulate the estimation problem of our concern. Let $\mathbf{x}^{(i)}$, $i = 1, \dots, N$ be independently and identically distributed

Table 1: Bias of d_i

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
$E(d_1)$	$E(d_2)$	$E(d_3)$	$E(d_4)$	$E(d_5)$	$E(d_6)$	$E(d_7)$	$E(d_8)$	$E(d_9)$	$E(d_{10})$
0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
0.216	0.172	0.142	0.118	0.097	0.080	0.064	0.050	0.037	0.024
0.120	0.120	0.120	0.120	0.120	0.080	0.080	0.080	0.080	0.080
0.225	0.176	0.143	0.117	0.095	0.077	0.061	0.047	0.035	0.023
0.140	0.140	0.140	0.140	0.140	0.060	0.060	0.060	0.060	0.060
0.245	0.189	0.149	0.116	0.090	0.069	0.053	0.040	0.029	0.019
0.160	0.160	0.160	0.160	0.160	0.040	0.040	0.040	0.040	0.040
0.272	0.206	0.159	0.121	0.087	0.053	0.039	0.029	0.021	0.014
0.180	0.180	0.180	0.180	0.180	0.020	0.020	0.020	0.020	0.020
0.300	0.226	0.173	0.129	0.090	0.028	0.020	0.015	0.011	0.007
0.198	0.198	0.198	0.198	0.198	0.002	0.002	0.002	0.002	0.002
0.328	0.245	0.187	0.139	0.094	0.003	0.002	0.002	0.001	0.001
0.200	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089
0.250	0.171	0.137	0.112	0.092	0.075	0.060	0.046	0.034	0.023
0.400	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067
0.419	0.134	0.106	0.087	0.071	0.058	0.046	0.036	0.026	0.018
0.600	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044
0.605	0.091	0.072	0.059	0.048	0.039	0.031	0.024	0.018	0.012
0.800	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
0.799	0.046	0.037	0.030	0.024	0.020	0.016	0.012	0.009	0.006
0.990	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
0.990	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.000	0.000

p -dimensional sample vectors with a covariance matrix $\boldsymbol{\Sigma}$. Suppose $N > p$. The sample covariance matrix

$$\mathbf{A} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}^{(i)} - \bar{\mathbf{x}})(\mathbf{x}^{(i)} - \bar{\mathbf{x}})',$$

where $\bar{\mathbf{x}}$ is a sample mean vector, is an unbiased estimator of $\boldsymbol{\Sigma}$. We consider the estimation problem of $\boldsymbol{\tau}$ (defined by (1)) based on \mathbf{A} .

We define the following notations;

$$\mathbf{X} = \begin{pmatrix} (\mathbf{x}^{(1)})' \\ \vdots \\ (\mathbf{x}^{(N)})' \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} (\mathbf{y}^{(1)})' \\ \vdots \\ (\mathbf{y}^{(N)})' \end{pmatrix}, \quad \mathbf{y}^{(i)} = \mathbf{x}^{(i)} - \bar{\mathbf{x}}, \quad i = 1, \dots, N,$$

$$\mathbf{C} = \mathbf{I}_N - N^{-1}\mathbf{1}\mathbf{1}',$$

where \mathbf{I}_N is the N -dimensional identity matrix, and $\mathbf{1}$ is the N -dimensional vector with unit as each element. We find that

$$\mathbf{Y} = \mathbf{C}\mathbf{X}.$$

The expression of \mathbf{A}

$$\mathbf{A} = \frac{1}{N-1} \mathbf{Y}'\mathbf{Y}$$

is inconvenient, since the rows of \mathbf{Y} are linearly constrained. Notice that

$$\mathbf{A} = \frac{1}{N-1} \mathbf{Y}'\mathbf{Y} = \frac{1}{N-1} \mathbf{X}'\mathbf{C}^2\mathbf{X} = \frac{1}{N-1} \mathbf{X}'\mathbf{C}\mathbf{X}.$$

Using the decomposition of $\mathbf{C} = \mathbf{O}_1\mathbf{O}_1'$, $\mathbf{O}_1 \in V_{N-1,N}$, where $V_{N-1,N}$ is the Stiefel Manifold, if we put $\mathbf{Z} = \mathbf{O}_1'\mathbf{X}$, then we have

$$\mathbf{A} = \frac{1}{N-1} \mathbf{Z}'\mathbf{Z}. \quad (6)$$

The distribution of \mathbf{A} is determined by \mathbf{Z} through (6), where \mathbf{Z} is not degenerated.

The most frequently postulated situation is that \mathbf{x}_i $i = 1, \dots, N$ is independently distributed as the p -variate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $(N-1)\mathbf{A}$ is distributed as Wishart distribution $W_p(N-1, \boldsymbol{\Sigma})$. This is distributionally equivalent to postulating that \mathbf{Z} is distributed as

$$\mathbf{Z} \sim N_{n \times p}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}), \quad n = N-1.$$

The density function of \mathbf{Z} is proportional to

$$\exp(-(1/2)\text{tr}\mathbf{Z}'\mathbf{Z}\boldsymbol{\Sigma}^{-1})|\boldsymbol{\Sigma}|^{-n/2}. \quad (7)$$

One of the natural generalizations of (7) is an elliptically contoured distribution, the density of which is given by

$$f(\text{tr}\mathbf{Z}'\mathbf{Z}\boldsymbol{\Sigma}^{-1})|\boldsymbol{\Sigma}|^{-n/2}. \quad (8)$$

We formulate our estimation problem as follows;

\mathbf{Z} is a $n \times p$ ($n \geq p \geq 2$) random matrix, and its density with respect to Lebesgue measure on R^{np} is given by (8) with some function $f(\cdot)$ on R , where $\boldsymbol{\Sigma}$ is an unknown positive definite p -dimensional matrix. We just observe

$$\mathbf{S} = \mathbf{Z}'\mathbf{Z}. \quad (9)$$

We consider the estimation of the population contribution rates $\boldsymbol{\tau}$ given by (1) based on \mathbf{S} .

2.2 Distribution of the Sample Contribution Rates

From (8) and (9), the density of \mathbf{S} is given by

$$c_1 f(\text{tr}\mathbf{S}\boldsymbol{\Sigma}^{-1})|\mathbf{S}|^{(n-p-1)/2}|\boldsymbol{\Sigma}|^{-n/2} \quad (10)$$

with some constant c_1 (the proof can be found in Appendix). When $f(x) = \exp(-x/2)$ this density function is that of a Wishart distribution. The distribution \mathbf{S} and the parameter $\boldsymbol{\Sigma}$ are equivariant with respect to the transformations $\mathbf{S} \rightarrow \mathbf{B}\mathbf{S}\mathbf{B}'$, $\boldsymbol{\Sigma} \rightarrow \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'$ for any p -dimensional nonsingular matrix \mathbf{B} .

The eigenvalues of \mathbf{S} are denoted by l_i , $1 \leq i \leq p$. We derive the distribution of the sample contribution rates

$$d_i = \frac{l_i}{\sum_{j=1}^p l_j}, \quad i = 1, \dots, p.$$

They are on the hyperplane

$$\mathcal{D} = \left\{ (d_1, \dots, d_p) \mid d_1 > \dots > d_p > 0, \sum_{i=1}^p d_i = 1 \right\}.$$

We use the coordinate system $\mathbf{d} = (d_1, \dots, d_{p-1})$ for \mathcal{D} . The range of \mathbf{d} is given by

$$\mathcal{D}^* = \left\{ \mathbf{d} \mid d_1 > \dots > d_{p-1} > 0, \sum_{i=1}^{p-1} d_i < 1 \right\}.$$

Theorem 1 *Let*

$$\mathbf{S} = \mathbf{H}\mathbf{L}\mathbf{H}', \quad \mathbf{L} = \text{diag}(l_1, \dots, l_p), \quad \mathbf{H} \in O(p) \quad (11)$$

be the spectral decomposition of \mathbf{S} , where $O(p)$ is the set of p -dimensional orthogonal matrices.

(i) The density function of (\mathbf{d}, \mathbf{H}) with respect to the product measure between Lebesgue measure on R^{p-1} and the invariant probability measure $\mu_{p,p}$ on $O(p)$ is given by

$$c_2 |\boldsymbol{\Sigma}|^{-n/2} F(\mathbf{d}) \left(\text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{D} \mathbf{H}' \right)^{-np/2}, \quad (12)$$

where c_2 is a constant, $\mathbf{D} = \text{diag}(d_1, \dots, d_{p-1}, 1 - \sum_{j=1}^{p-1} d_j)$, and $F(\cdot)$ is a positive-valued function on \mathcal{D}^* that is independent of $f(\cdot)$ in (8) or $\boldsymbol{\Sigma}$.

(ii) The density function of \mathbf{d} with respect to Lebesgue measure on R^{p-1} is given by

$$c_2 \prod_{i=1}^p \tau_i^{-n/2} F(\mathbf{d}) \int_{O(p)} \left(\text{tr} \mathbf{T}^{-1} \mathbf{H} \mathbf{D} \mathbf{H}' \right)^{-np/2} \mu_{p,p}(d\mathbf{H}), \quad (13)$$

where $\mathbf{T} = \text{diag}(\tau_1, \dots, \tau_p)$.

Proof From (10), the density function of $\mathbf{l} = (l_1, \dots, l_p)$ and \mathbf{H} with respect to the product measure between Lebesgue measure on R^p and $\mu_{p,p}$ on $O(p)$ is given by

$$\tilde{c} c_1 |\boldsymbol{\Sigma}|^{-n/2} \prod_{i=1}^p l_i^{(n-p-1)/2} \prod_{i < j} (l_i - l_j) f \left(\text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{L} \mathbf{H}' \right)$$

with some positive constant \tilde{c} . (See e.g. (22) on the p105 of Muirhead(1982).)

Let $t = \sum_{i=1}^p l_i$ and consider the transformation of coordinates

$$\mathbf{l} \rightarrow (t, \mathbf{d}).$$

Since the Jacorbian is given by $J(\mathbf{l} \rightarrow (t, \mathbf{d})) = t^{p-1}$, the density function of t, \mathbf{d} and \mathbf{H} with respect to $d\mathbf{t} \times d\mathbf{d} \times \mu_{p,p}(d\mathbf{H})$ is given by

$$\tilde{c} c_1 |\boldsymbol{\Sigma}|^{-n/2} t^{np/2-1} F(\mathbf{d}) f \left(t \text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{D} \mathbf{H}' \right), \quad (14)$$

where

$$\begin{aligned} F(\mathbf{d}) &= \left(\prod_{1 \leq i < j \leq p-1} (d_i - d_j) \right) \left(\prod_{i=1}^{p-1} (d_i - (1 - \sum_{j=1}^{p-1} d_j)) \right) \\ &\quad \times \left(\prod_{i=1}^{p-1} d_i^{(n-p-1)/2} \right) \left(1 - \sum_{j=1}^{p-1} d_j \right)^{(n-p-1)/2}. \end{aligned}$$

Integrate (14) over $\{t \mid 0 < t < \infty\}$, then

$$\begin{aligned} &\tilde{c} c_1 |\boldsymbol{\Sigma}|^{-n/2} F(\mathbf{d}) \int_0^\infty t^{np/2-1} f(ta) dt \quad \left(a = \text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{D} \mathbf{H}' > 0 \right) \\ &= \tilde{c} c_1 |\boldsymbol{\Sigma}|^{-n/2} F(\mathbf{d}) a^{-np/2} \int_0^\infty t^{np/2-1} f(t) dt. \end{aligned}$$

Now we have the density function of (\mathbf{d}, \mathbf{H}) with respect to $d\mathbf{d} \times \mu_{p,p}(d\mathbf{H})$ as

$$\tilde{c} c_1 |\boldsymbol{\Sigma}|^{-n/2} \left(\int_0^\infty t^{np/2-1} f(t) dt \right) F(\mathbf{d}) \left(\text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{D} \mathbf{H}' \right)^{-np/2}.$$

Substituting

$$\tilde{c} c_1 \int_0^\infty t^{np/2-1} f(t) dt$$

with c_2 , we have (12).

Integrating (12) over $O(p)$, we have the density of \mathbf{d} as

$$c_2 \prod_{i=1}^p \lambda_i^{-n/2} F(\mathbf{d}) \int_{O(p)} \left(\text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{D} \mathbf{H}' \right)^{-np/2} \mu_{p,p}(d\mathbf{H}). \quad (15)$$

Let

$$\boldsymbol{\Sigma} = \tilde{\mathbf{H}}' \boldsymbol{\Lambda} \tilde{\mathbf{H}}, \quad \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p), \quad \tilde{\mathbf{H}} \in O(p)$$

be the spectral decomposition of $\boldsymbol{\Sigma}$. Since $\mu_{p,p}$ is the invariant probability on $O(p)$, $\tilde{\mathbf{H}} \mathbf{H}$ has the same distribution as \mathbf{H} . Therefore (15) equals

$$c_2 \prod_{i=1}^p \lambda_i^{-n/2} F(\mathbf{d}) \int_{O(p)} \left(\text{tr} \boldsymbol{\Lambda}^{-1} \mathbf{H} \mathbf{D} \mathbf{H}' \right)^{-np/2} \mu_{p,p}(d\mathbf{H}).$$

Substituting λ_i with $\tau_i \sum_{j=1}^p \lambda_j$, we have (13). ■

It is noteworthy that the distribution of \mathbf{d} is independent of $f(\cdot)$ in (8) and depends on $\boldsymbol{\Sigma}$ only through the population contribution rates $\boldsymbol{\tau}$. Note that Johnson and Grayvill (1972) deals with the distribution of the sample contribution rates when $\boldsymbol{\Sigma} = I_p$.

2.3 New Estimator

In order to derive a new estimator $\hat{\boldsymbol{\tau}}(\mathbf{d}) = (\hat{\tau}_1(\mathbf{d}), \dots, \hat{\tau}_p(\mathbf{d}))$ of $\boldsymbol{\tau} = (\tau_1, \dots, \tau_p)$ that has a certain superiority to the classical estimator, we take a decision theoretic approach here, that is, we compare estimators via their risks with respect to a certain loss function. Straightforward approach is to use a loss function that directly measures the distance between $\boldsymbol{\tau}$ and its estimator $\hat{\boldsymbol{\tau}}$. However the exact property of the sample contribution rates \mathbf{d} are difficult to derive under the assumption of small samples. Instead we evaluate the performance of $\hat{\boldsymbol{\tau}}(\mathbf{d})$ as the components of an estimator of $\boldsymbol{\Sigma}$.

Consider an estimator of $\boldsymbol{\Sigma}$ combining $\hat{\boldsymbol{\tau}}$ and the sample eigenvectors \mathbf{H} in (11) as follows;

$$\hat{\boldsymbol{\Sigma}} = \mathbf{H} \hat{\mathbf{T}}(\mathbf{d}) \mathbf{H}', \quad \hat{\mathbf{T}}(\mathbf{d}) = \text{diag}(\hat{\tau}_1(\mathbf{d}), \dots, \hat{\tau}_p(\mathbf{d})). \quad (16)$$

Our approach is based on the following observations. According to Theorem 1, the distribution of \mathbf{d} is determined by $\boldsymbol{\tau}$. Therefore we can suppose that

$\text{tr}\boldsymbol{\Sigma} = \sum_{i=1}^p \lambda_i = 1$ without loss of generality. In this case, the population contribution rates are equal to the population eigenvalues. In addition, the sample eigenvectors \mathbf{H} are M.L.E., hence consistent under the large sample asymptotics if $f(\cdot)$ in (10) is monotonically decreasing. (See Paul (2007) and Nadler (2008) for the discrepancy between the sample eigenvectors and the population counterparts for a large-dimensional matrix.) Therefore (16) is supposed to be a good estimator of $\boldsymbol{\Sigma}$ if $\hat{\boldsymbol{\tau}}$ is a good estimator of $\boldsymbol{\tau}$.

The most common loss function about $\boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\Sigma}}$ is the entropy loss function (Stein's loss function)

$$L(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \text{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log(|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}|) - p. \quad (17)$$

We evaluate the performance of $\hat{\boldsymbol{\Sigma}}$ through its risk with respect to this loss function.

We consider one class of simple estimators given by

$$\hat{\boldsymbol{\tau}}^* = (\hat{\tau}_1^*, \dots, \hat{\tau}_p^*), \quad \hat{\tau}_i^* = \beta_i^* d_i, \quad 1 \leq i \leq p, \quad (18)$$

where β_i^* , $i = 1, \dots, p$ are positive constants. The classical estimator denoted by $\hat{\boldsymbol{\tau}}^0$ is given by

$$\hat{\boldsymbol{\tau}}^0 = (\hat{\tau}_1^0, \dots, \hat{\tau}_p^0), \quad \hat{\tau}_i^0 = \beta_i^0 d_i, \quad 1 \leq i \leq p, \quad (19)$$

where $\beta_i^0 = 1$, $i = 1, \dots, p$. Correspondingly we define the two estimators $\hat{\boldsymbol{\Sigma}}^*$ and $\hat{\boldsymbol{\Sigma}}^0$ as follows;

$$\hat{\boldsymbol{\Sigma}}^* = \mathbf{H}\hat{\mathbf{T}}^*\mathbf{H}', \quad \hat{\mathbf{T}}^* = \text{diag}(\hat{\tau}_1^*, \dots, \hat{\tau}_p^*), \quad (20)$$

$$\hat{\boldsymbol{\Sigma}}^0 = \mathbf{H}\hat{\mathbf{T}}^0\mathbf{H}', \quad \hat{\mathbf{T}}^0 = \text{diag}(\hat{\tau}_1^0, \dots, \hat{\tau}_p^0). \quad (21)$$

We have the following result on the superiority of $\hat{\boldsymbol{\Sigma}}^*$ to $\hat{\boldsymbol{\Sigma}}^0$.

Theorem 2 *If β_i^* ($i = 1, \dots, p$) satisfy the following three conditions, then $\hat{\boldsymbol{\Sigma}}^*$ dominates $\hat{\boldsymbol{\Sigma}}^0$ with respect to the loss function (17).*

For some m ($1 \leq m \leq p - 1$), the next two inequalities hold;

$$0 < \beta_1^* \leq \dots \leq \beta_m^* \leq 1 \leq \beta_{m+1}^* \leq \dots \leq \beta_p^*, \quad (22)$$

$$\sum_{i=1}^m (n + p - 1 - 2i)(\beta_i^* - 1) + \sum_{i=m+1}^p (n + p + 1 - 2i)(\beta_i^* - 1) \leq 0. \quad (23)$$

Moreover, the third inequality

$$\sum_{i=1}^p (\beta_i^*)^{-1} \leq p \quad (24)$$

holds.

Proof Since both $\hat{\Sigma}^*$ and $\hat{\Sigma}^0$ is the function of \mathbf{d} and \mathbf{H} , their distributions are independent of $f(\cdot)$ from the result (i) of Theorem 1. Therefore we can suppose $f(x) = \exp(-x/2)$, that is, \mathbf{S} is distributed as a Wishart matrix;

$$\mathbf{S} \sim W_p(n, \Sigma). \quad (25)$$

If \mathbf{S} is distributed as in (25), the following Stein-Haff identity holds. (Exactly speaking, it is the application of Stein-Haff identity to an orthogonally equivariant estimator, see e.g. Lemma 2.1 of Dey and Srinivasan (1986)); Suppose \mathbf{S} is decomposed as in (11) and $\hat{\Sigma}$ is given by

$$\hat{\Sigma} = \mathbf{H} \text{diag}(\phi_1(\mathbf{l}), \dots, \phi_p(\mathbf{l})) \mathbf{H}', \quad \mathbf{l} = (l_1, \dots, l_p).$$

Then

$$E[\text{tr}(\hat{\Sigma} \Sigma^{-1})] = E[G(\hat{\Sigma}, \mathbf{l})],$$

where

$$G(\hat{\Sigma}, \mathbf{l}) = 2 \sum_{1 \leq i < j \leq p} \frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j} + 2 \sum_{i=1}^p \frac{\partial \phi_i(\mathbf{l})}{\partial l_i} + (n - p - 1) \sum_{i=1}^p \frac{\phi_i(\mathbf{l})}{l_i}. \quad (26)$$

If we use this identity, we have the following equation.

$$\begin{aligned} & E[L(\hat{\Sigma}^*, \Sigma)] - E[L(\hat{\Sigma}^0, \Sigma)] \\ &= E[\text{tr}(\hat{\Sigma}^* \Sigma^{-1}) - \log |\hat{\Sigma}^*|] - E[\text{tr}(\hat{\Sigma}^0 \Sigma^{-1}) - \log |\hat{\Sigma}^0|] \\ &= E[G(\hat{\Sigma}^*, \mathbf{l}) - G(\hat{\Sigma}^0, \mathbf{l}) - \log |\hat{\Sigma}^*| + \log |\hat{\Sigma}^0|]. \end{aligned} \quad (27)$$

Substituting (18) and (20) into (26), followed by simple calculation, we have

$$\begin{aligned} G(\hat{\Sigma}^*, \mathbf{l}) &= \frac{2}{\sum_{j=1}^p l_j} \sum_{1 \leq i < j \leq p} \frac{\beta_i^* l_i - \beta_j^* l_j}{l_i - l_j} + \frac{2}{\sum_{j=1}^p l_j} \sum_{i=1}^p \beta_i^* \frac{\sum_{j=1}^p l_j - l_i}{\sum_{j=1}^p l_j} \\ &\quad + \frac{n - p - 1}{\sum_{j=1}^p l_j} \sum_{i=1}^p \beta_i^* \\ &= \frac{1}{\sum_{j=1}^p l_j} \left\{ 2 \sum_{1 \leq i < j \leq p} \frac{\beta_i^* l_i - \beta_i^* l_j + \beta_j^* l_j - \beta_j^* l_i}{l_i - l_j} \right. \\ &\quad \left. + 2 \sum_{i=1}^p \beta_i^* (1 - d_i) + (n - p - 1) \sum_{i=1}^p \beta_i^* \right\} \\ &= \frac{1}{\sum_{j=1}^p l_j} \left\{ 2 \sum_{1 \leq i < j \leq p} (\beta_i^* - \beta_j^*) \frac{l_j}{l_i - l_j} + 2 \sum_{i=1}^p (p - i) \beta_i^* \right. \\ &\quad \left. + 2 \sum_{i=1}^p \beta_i^* (1 - d_i) + (n - p - 1) \sum_{i=1}^p \beta_i^* \right\} \\ &= \frac{1}{\sum_{j=1}^p l_j} \left\{ 2 \sum_{1 \leq i < j \leq p} (\beta_i^* - \beta_j^*) \frac{l_j}{l_i - l_j} + \sum_{i=1}^p (n + p + 1 - 2i - 2d_i) \beta_i^* \right\} \end{aligned} \quad (28)$$

If we substitute β_i^* in (28) with $\beta_i^0 = 1$ ($1 \leq i \leq p$), we have $G(\hat{\Sigma}^0, \mathbf{l})$. From these results, the inside of the brackets of the right-hand side in (27) turns out to be

$$\begin{aligned} & \left(\sum_{j=1}^p l_j \right)^{-1} \left\{ 2 \sum_{1 \leq i < j \leq p} (\beta_i^* - \beta_j^*) \frac{l_j}{l_i - l_j} + \sum_{i=1}^p (n + p + 1 - 2i - 2d_i)(\beta_i^* - 1) \right\} \\ & - \log \prod_{i=1}^p \beta_i^*. \end{aligned} \quad (29)$$

Since $\beta_i^* \leq \beta_j^*$ for $1 \leq i < j \leq p$ from (22), (29) is less than or equal to

$$\left(\sum_{j=1}^p l_j \right)^{-1} \sum_{i=1}^p (n + p + 1 - 2i - 2d_i)(\beta_i^* - 1) - \log \prod_{i=1}^p \beta_i^*. \quad (30)$$

(22) says that $\beta_i^* \leq 1$ if $1 \leq i \leq m$ and that $\beta_i^* \geq 1$ if $m + 1 \leq i \leq p$. Using this fact together with the inequality $0 \leq d_i \leq 1$, $\forall i$, we notice that (30) is less than or equal to

$$\left(\sum_{j=1}^p l_j \right)^{-1} \left\{ \sum_{i=1}^m (n + p - 1 - 2i)(\beta_i^* - 1) + \sum_{i=m+1}^p (n + p + 1 - 2i)(\beta_i^* - 1) \right\} - \log \prod_{i=1}^p \beta_i^*.$$

From (23), this is less than or equal to $\sum_{i=1}^p \log(\beta_i^*)^{-1}$. Because of the inequality $\log(x + 1) \leq x$, $\forall x > -1$, we have

$$\sum_{i=1}^p \log(\beta_i^*)^{-1} = \sum_{i=1}^p \log\{(\beta_i^*)^{-1} - 1 + 1\} \leq \sum_{i=1}^p \{(\beta_i^*)^{-1} - 1\},$$

which is nonpositive by (24). ■

(22) of Theorem 2 means $\boldsymbol{\tau}^*$ modifies the bias of the classical estimator which we mentioned in Section 1, since lighter weight is given to d_i for the smaller i 's and heavier weight for the larger i 's.

Choose an integer q such that $1 \leq q \leq p/2 - 1$. Let $\beta_i^{(q)}$ ($i = 1, \dots, p$) be defined as

$$\beta_i^{(q)} = \begin{cases} n(n + p - 2q + 1 - 2i)^{-1} & \text{if } 1 \leq i \leq m - q, \\ 1 & \text{if } m - q + 1 \leq i \leq p - m + q, \\ n(n + p + 2q + 1 - 2i)^{-1} & \text{if } p - m + q + 1 \leq i \leq p, \end{cases} \quad (31)$$

where $m = [p/2]$, i.e., the largest integer that does not exceed $p/2$. Then $\beta_i^{(q)}$ ($i = 1, \dots, p$) satisfy (22), (23) and (24). In fact, (22) is clearly satisfied

from the definition. (23) and (24) are also satisfied as follows;

$$\begin{aligned}
& \sum_{i=1}^m (n+p-1-2i)(\beta_i^{(q)} - 1) + \sum_{i=m+1}^p (n+p+1-2i)(\beta_i^{(q)} - 1) \\
&= \sum_{i=1}^{m-q} (n+p-1-2i)(\beta_i^{(q)} - 1) + \sum_{i=p-m+q+1}^p (n+p+1-2i)(\beta_i^{(q)} - 1) \\
&\leq \sum_{i=1}^{m-q} (n+p-2q+1-2i)(\beta_i^{(q)} - 1) + \sum_{i=p-m+q+1}^p (n+p+2q+1-2i)(\beta_i^{(q)} - 1) \\
&= \sum_{i=1}^{m-q} (-p+2q-1+2i) + \sum_{i=p-m+q+1}^p (-p-2q-1+2i) \\
&= (m-q)\{(-p+2q-1) + (m-q+1) + (-p-2q-1) + (2p-m+q+1)\} \\
&= 0,
\end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^p (\beta_i^{(q)})^{-1} - p \\
&= \sum_{i=1}^p \{(\beta_i^{(q)})^{-1} - 1\} \\
&= n^{-1} \left\{ \sum_{i=1}^{m-q} (p-2q+1-2i) + \sum_{i=p-m+q+1}^p (p+2q+1-2i) \right\} \\
&= 0.
\end{aligned}$$

We give two examples of the estimators that satisfy the three conditions in Theorem 2. Let $q = 1$, then

$$\beta_i^{(1)} = \begin{cases} n(n+p-1-2i)^{-1} & \text{if } 1 \leq i \leq m-1, \\ 1 & \text{if } m \leq i \leq p-m+1, \\ n(n+p+3-2i)^{-1} & \text{if } p-m+2 \leq i \leq p, \end{cases} \quad (32)$$

The specific value of $\beta_i^{(1)}$ ($1 \leq i \leq p$) is given as follows;
if p is even

$$\begin{array}{c|c|c|c|c|c|c|c|c|c}
\beta_1^{(1)} & \beta_2^{(1)} & \cdots & \beta_{m-1}^{(1)} & \beta_m^{(1)} & \beta_{m+1}^{(1)} & \beta_{m+2}^{(1)} & \cdots & \beta_{p-1}^{(1)} & \beta_p^{(1)} \\
\hline
\frac{n}{n+p-3} & \frac{n}{n+p-5} & \cdots & \frac{n}{n+1} & 1 & 1 & \frac{n}{n-1} & \cdots & \frac{n}{n-p+5} & \frac{n}{n-p+3}
\end{array}$$

if p is odd

$$\begin{array}{c|c|c|c|c|c|c|c|c|c|c}
\beta_1^{(1)} & \beta_2^{(1)} & \cdots & \beta_{m-1}^{(1)} & \beta_m^{(1)} & \beta_{m+1}^{(1)} & \beta_{m+2}^{(1)} & \beta_{m+3}^{(1)} & \cdots & \beta_{p-1}^{(1)} & \beta_p^{(1)} \\
\hline
\frac{n}{n+p-3} & \frac{n}{n+p-5} & \cdots & \frac{n}{n+2} & 1 & 1 & 1 & \frac{n}{n-2} & \cdots & \frac{n}{n-p+5} & \frac{n}{n-p+3}
\end{array}$$

The estimator

$$\hat{\Sigma}^{(1)} = \mathbf{H}\hat{\mathbf{T}}^{(1)}\mathbf{H}', \quad \hat{\mathbf{T}}^{(1)} = \text{diag}(\hat{\tau}_1^{(1)}, \dots, \hat{\tau}_p^{(1)}), \quad (33)$$

where $\hat{\tau}_i^{(1)} = \beta_i^{(1)}d_i$, $1 \leq i \leq p$, dominates $\hat{\Sigma}^0$ if $p \geq 4$. Another estimator that satisfies the three conditions of Theorem 2 is given by $q = 2$, which leads to

$$\beta_i^{(2)} = \begin{cases} n(n+p-3-2i)^{-1} & \text{if } 1 \leq i \leq m-2, \\ 1 & \text{if } m-1 \leq i \leq p-m+2, \\ n(n+p+5-2i)^{-1} & \text{if } p-m+3 \leq i \leq p. \end{cases} \quad (34)$$

The specific value of $\beta_i^{(2)}$ ($1 \leq i \leq p$) is given as follows; if p is even

$\beta_1^{(2)}$	$\beta_2^{(2)}$	\dots	$\beta_{m-2}^{(2)}$	$\beta_{m-1}^{(2)}$	\dots	$\beta_{m+2}^{(2)}$	$\beta_{m+3}^{(2)}$	\dots	$\beta_{p-1}^{(2)}$	$\beta_p^{(2)}$
$\frac{n}{n+p-5}$	$\frac{n}{n+p-7}$	\dots	$\frac{n}{n+1}$	1	\dots	1	$\frac{n}{n-1}$	\dots	$\frac{n}{n-p+7}$	$\frac{n}{n-p+5}$

if p is odd

$\beta_1^{(2)}$	$\beta_2^{(2)}$	\dots	$\beta_{m-2}^{(1)}$	$\beta_{m-1}^{(2)}$	\dots	$\beta_{m+3}^{(2)}$	$\beta_{m+4}^{(2)}$	\dots	$\beta_{p-1}^{(2)}$	$\beta_p^{(2)}$
$\frac{n}{n+p-5}$	$\frac{n}{n+p-7}$	\dots	$\frac{n}{n+2}$	1	\dots	1	$\frac{n}{n-2}$	\dots	$\frac{n}{n-p+7}$	$\frac{n}{n-p+5}$

Note that

$$\begin{aligned} \beta_i^{(1)} &\leq \beta_i^{(2)} \leq \beta_i^{(0)} (\equiv 1) \text{ for } 1 \leq i \leq m, \\ \beta_i^{(1)} &\geq \beta_i^{(2)} \geq \beta_i^{(0)} (\equiv 1) \text{ for } m+1 \leq i \leq p. \end{aligned} \quad (35)$$

The estimator

$$\hat{\Sigma}^{(2)} = \mathbf{H}\hat{\mathbf{T}}^{(2)}\mathbf{H}', \quad \hat{\mathbf{T}}^{(2)} = \text{diag}(\hat{\tau}_1^{(2)}, \dots, \hat{\tau}_p^{(2)}), \quad (36)$$

where $\hat{\tau}_i^{(2)} = \beta_i^{(2)}d_i$, $1 \leq i \leq p$, dominates $\hat{\Sigma}^0$ if $p \geq 6$.

3 Simulation Study

In this section, we examine by simulation other preferable properties of the new estimator $\hat{\tau}^* = (\hat{\tau}_1^*, \dots, \hat{\tau}_p^*)$,

$$\hat{\tau}_i^* = \beta_i^*d_i, \quad 1 \leq i \leq p,$$

where β_i^* 's satisfy (22)–(24), especially when $\beta_i^* = \beta_i^{(1)}$ or $\beta_i^{(2)}$, $1 \leq i \leq p$. We use the notation $\hat{\tau}^{(j)} = (\hat{\tau}_1^{(j)}, \dots, \hat{\tau}_p^{(j)})$,

$$\hat{\tau}_i^{(j)} = \beta_i^{(j)}d_i, \quad j = 1, 2, \quad 1 \leq i \leq p.$$

Table 2: Risk w.r.t. Quadratic Loss

τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8	τ_9	τ_{10}	$\hat{\tau}^{(0)}$	$\hat{\tau}^{(1)}$	$\hat{\tau}^{(2)}$
0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	3.57	2.11	2.56
0.11	0.11	0.11	0.11	0.11	0.09	0.09	0.09	0.09	0.09	2.71	1.57	1.91
0.12	0.12	0.12	0.12	0.12	0.08	0.08	0.08	0.08	0.08	2.20	1.25	1.53
0.13	0.13	0.13	0.13	0.13	0.07	0.07	0.07	0.07	0.07	1.96	1.13	1.36
0.14	0.14	0.14	0.14	0.14	0.06	0.06	0.06	0.06	0.06	1.85	1.10	1.31
0.15	0.15	0.15	0.15	0.15	0.05	0.05	0.05	0.05	0.05	1.84	1.14	1.32
0.16	0.16	0.16	0.16	0.16	0.04	0.04	0.04	0.04	0.04	1.87	1.21	1.38
0.17	0.17	0.17	0.17	0.17	0.03	0.03	0.03	0.03	0.03	1.89	1.27	1.42
0.18	0.18	0.18	0.18	0.18	0.02	0.02	0.02	0.02	0.02	1.91	1.32	1.46
0.19	0.19	0.19	0.19	0.19	0.01	0.01	0.01	0.01	0.01	1.94	1.37	1.50
0.20	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	2.71	1.78	2.15
0.30	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	2.86	1.92	2.31
0.40	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	2.97	2.02	2.42
0.50	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	3.09	2.12	2.53
0.60	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	3.24	2.23	2.66
0.70	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	3.44	2.39	2.84
0.80	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	3.65	2.56	3.03
0.90	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	3.94	2.79	3.28

3.1 Risk Comparison

We will compare the estimator $\hat{\tau}^{(1)}$ and $\hat{\tau}^{(2)}$ with the classical estimator $\hat{\tau}^{(0)}$ through their risks with respect to the quadratic loss function;

$$QL(\hat{\tau}, \tau) = \sum_{i=1}^p (1 - \hat{\tau}_i/\tau_i)^2. \quad (37)$$

According to Theorem 2, the plug-in estimator $\hat{\Sigma}^{(j)}$ made from $\hat{\tau}^{(j)}$ $j = 1, 2$ dominates another plug-in estimator $\hat{\Sigma}^{(0)}$ from $\hat{\tau}^{(0)}$ with respect to the entropy loss function. We are interested in a more direct comparison among $\hat{\tau}^{(0)}$, $\hat{\tau}^{(1)}$ and $\hat{\tau}^{(2)}$ using (37).

We generated 10000 random 10-dimensional Wishart matrices with the degree of freedom 30 under several patterns of the population contribution rates, $\tau = (\tau_1, \dots, \tau_{10})$. The Table 2 shows the simulation result, where the first 10 numbers in each row are the population contribution rates and the last three numbers are the simulated risks for the three estimators $\hat{\tau}^{(j)}$, $j = 0, 1, 2$ (all the numbers are rounded to the second decimal place). The risk of $\hat{\tau}^{(1)}$ is smaller than that of $\hat{\tau}^{(0)}$ by 30% to 40%. Since $\hat{\tau}^{(2)}$ is located between $\hat{\tau}^{(0)}$ and $\hat{\tau}^{(1)}$ (see (35)), its risk reduction is smaller than $\hat{\tau}^{(1)}$. Nevertheless it still reduces the risk by 17% to 30% compared to $\hat{\tau}^{(0)}$. From these results, we can conclude that the new estimators are substantially improved over the classical estimator in view of the quadratic risk.

3.2 Estimation of Dimension

As we mentioned in Section 1, $\boldsymbol{\tau}$, the population contribution rates, is one of the most basic tools for deciding the dimension in principal component analysis or factor analysis. As the first step in deciding the dimension, the choice of an estimator $\hat{\boldsymbol{\tau}}$ for $\boldsymbol{\tau}$ is an important task, hence we are interested in how the new estimator, $\hat{\boldsymbol{\tau}}^{(*)}$, makes a difference compared to the classical estimator, $\hat{\boldsymbol{\tau}}^{(0)}$, in the decision of the dimension.

Suppose that $\boldsymbol{x} = (x_1, \dots, x_p)'$ are generated in the following m -factor model ;

$$\boldsymbol{x} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{z} + \boldsymbol{e}, \quad (38)$$

where \boldsymbol{a} is a constant p -dimensional vector, \boldsymbol{B} is a $p \times m$ ($p \geq m$) factor loading constant matrix with the rank of m , \boldsymbol{z} is a m -dimensional random factor, and \boldsymbol{e} is the p -dimensional error term which is independent of \boldsymbol{z} . If we suppose the covariance matrices of \boldsymbol{z} and \boldsymbol{e} are respectively given by

$$V(\boldsymbol{z}) = \boldsymbol{\Sigma}_0, \quad V(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_p,$$

then the covariance matrix $\boldsymbol{\Sigma}$ of \boldsymbol{x} equals

$$\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Sigma}_0\boldsymbol{B}' + \sigma^2 \boldsymbol{I}_p. \quad (39)$$

If we denote the eigenvalues of $\boldsymbol{B}\boldsymbol{\Sigma}_0\boldsymbol{B}'$ by ξ_i ($1 \leq i \leq p$), then λ_i ($1 \leq i \leq p$), the eigenvalues of $\boldsymbol{\Sigma}$, are given by

$$\lambda_i = \begin{cases} \xi_i + \sigma^2 & \text{if } i = 1, \dots, m, \\ \sigma^2 & \text{if } i = m + 1, \dots, p, \end{cases} \quad (40)$$

since $\xi_i = 0$, ($m + 1 \leq i \leq p$). In the large-dimensional random matrix theory, the model (38) is called ‘‘spiked covariance model’’, since the part of $\xi_i + \sigma^2$ seems spiked into the long flat part σ^2 . If \boldsymbol{z} and \boldsymbol{e} have normal distributions, then \boldsymbol{x} is also normally distributed with the covariance matrix (39), hence we can suppose that $\boldsymbol{S} \sim W_p(n, \boldsymbol{\Sigma})$.

We made a simulation under the condition

$$\boldsymbol{S} \sim W_{10}(n, \boldsymbol{\Lambda}), \quad n = 30, 100, \quad \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p),$$

where λ_i ($1 \leq i \leq p$) is given by (40) with ten patterns of (m, ξ_1, \dots, ξ_m) and σ^2 fixed to be unit. \boldsymbol{S} is generated 10000 times, and for each time we recorded the dimension decided from the six methods composed by the combination of two criterions in Section1 and three estimators $\hat{\boldsymbol{\tau}}^{(0)}$, $\hat{\boldsymbol{\tau}}^{(1)}$, $\hat{\boldsymbol{\tau}}^{(2)}$; the two criterions are ‘‘the cumulative percentage of the (estimated) population eigenvalues’’ with $t^* = 0.8$ in (2) (say criterion 1) and ‘‘the relative size of each (estimated) population eigenvalue’’ (say criterion 2).

Table 3 and 4 are the result of the simulation. To explain the meaning of each number in the table, take the first case in Table 3 as an example,

Table 3: Histogram of Estimated Dimension, n=30

Case 1	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	19	19	19	19	19	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	104	9895	1	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	9	9991	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	0	339	9661	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	0	285	6480	3235	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	0	420	6346	3234	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	0	2275	4930	2795	0	0	0	0	0
Case 2	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	23.5	23.5	23.5	23.5	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	7348	2652	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	9828	172	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	17	9982	1	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	1	1087	8912	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	5	1348	8647	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	2	3801	6197	0	0	0	0	0	0
Case 3	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	31	31	31	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	376	9624	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	2410	6641	948	1	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	9987	13	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	22	9978	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	50	9950	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	30	9970	0	0	0	0	0	0	0
Case 4	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	46	46	1	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	10000	0	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	1	137	3545	5907	410	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	8008	1903	88	1	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	10000	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	10000	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	10000	0	0	0	0	0	0	0	0
Case 5	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	91	1	1	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	9991	9	0	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	2	776	8170	1052	0	0	0
C.1 & $\hat{\tau}^{(2)}$	1184	3743	3916	1064	93	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	10000	0	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	10000	0	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	10000	0	0	0	0	0	0	0	0	0

Case 6	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	27	23	19	15	11	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	1018	8981	1	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	48	9952	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	0	1535	8465	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	2	1483	7619	896	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	12	1922	7171	895	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	2	5407	3904	687	0	0	0	0	0
Case 7	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	32	25.5	19	12.5	6	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	5931	4069	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	688	9311	1	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	2	6764	3234	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	35	4702	5254	9	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	107	5209	4676	8	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	49	8246	1701	4	0	0	0	0	0
Case 8	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	37	28	19	10	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	36	9847	117	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	9013	987	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	910	9090	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	107	7496	2397	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	317	7705	1978	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	168	9347	485	0	0	0	0	0	0
Case 9	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	46	31	16	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	4471	5529	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	598	6457	2923	22	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	9971	29	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	1022	8978	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	1756	8244	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	1234	8766	0	0	0	0	0	0	0
Case 10	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	81	11	1	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	6232	3767	1	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	14	3144	6766	76	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	4482	4664	824	30	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	3939	6061	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	5952	4048	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	5171	4829	0	0	0	0	0	0	0	0

Table 4: Histogram of Estimated Case, n=100

Case 1	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	19	19	19	19	19	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	0	9536	464	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	1259	8741	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	0	1922	8078	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	0	0	50	9950	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	0	0	50	9950	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	0	0	50	9950	0	0	0	0	0
Case 2	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	23.5	23.5	23.5	23.5	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	287	9713	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	10000	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	9	9991	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	0	1	9999	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	0	1	9999	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	0	2	9998	0	0	0	0	0	0
Case 3	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	31	31	31	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	10000	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	10000	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	10000	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	0	10000	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	0	10000	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	0	10000	0	0	0	0	0	0	0
Case 4	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	46	46	1	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	10000	0	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	10000	0	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	10000	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	10000	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	10000	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	10000	0	0	0	0	0	0	0	0
Case 5	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	91	1	1	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	10000	0	0	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	9995	5	0	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	9999	1	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	10000	0	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	10000	0	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	10000	0	0	0	0	0	0	0	0	0

Case 6	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	27	23	19	15	11	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	0	9998	2	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	0	8849	1151	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	0	9422	578	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	0	15	5304	4681	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	0	20	5299	4681	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	0	87	5239	4674	0	0	0	0	0
Case 7	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	32	25.5	19	12.5	6	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	1113	8887	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	1	9999	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	25	9975	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	0	1095	8905	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	0	1228	8772	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	0	2321	7679	0	0	0	0	0	0
Case 8	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	37	28	19	10	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	0	9964	36	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	5754	4246	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	0	8784	1216	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	0	6280	3720	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	0	6531	3469	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	0	7859	2141	0	0	0	0	0	0
Case 9	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	46	31	16	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	0	1628	8372	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	0	0	10000	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	0	49	9951	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	0	19	9981	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	0	38	9962	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	0	22	9978	0	0	0	0	0	0	0
Case 10	1	2	3	4	5	6	7	8	9	10
$\lambda_1, \dots, \lambda_{10}$	81	11	1	1	1	1	1	1	1	1
C.1 & $\hat{\tau}^{(0)}$	6679	3321	0	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(1)}$	208	9792	0	0	0	0	0	0	0	0
C.1 & $\hat{\tau}^{(2)}$	1108	8892	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(0)}$	2952	7048	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(1)}$	4039	5961	0	0	0	0	0	0	0	0
C.2 & $\hat{\tau}^{(2)}$	3612	6388	0	0	0	0	0	0	0	0

where λ_i , $i = 1, \dots, 10$ are given by 19, 19, 19, 19, 19, 1, 1, 1, 1, 1. (Note that for every case, λ_i 's are designed so that $\sum_{i=1}^p \lambda_i = 100$, hence each λ_i equals τ_i in percent figures.) The row of "C.1 & $\hat{\tau}^{(0)}$ " is the histogram of the estimated dimension by the combination of criterion 1 and $\hat{\tau}^{(0)}$. The boldface position (in this case "1") indicates that the "true" dimension decided by criterion 1 from the population eigenvalues, λ_i 's. In this simulation, the "true" dimensions are designed to take the same value by either criterion 1 or 2.

We can observe following points from Table 3. With respect to the criterion 1, we notice that the classical estimator, $\hat{\tau}^{(0)}$, tends to underestimate the dimension (see Case 1, 2, 7, 9), while $\hat{\tau}^{(1)}$ tend to overestimate it (see Case 3–10). From (35), we notice that $\hat{\tau}^{(2)}$ is located between $\hat{\tau}^{(0)}$ and $\hat{\tau}^{(1)}$. Though $\hat{\tau}^{(2)}$ is still likely to overestimate the dimension (see Case 5, 6, 8, 10), the tendency is weakened compared to $\hat{\tau}^{(1)}$. On the criterion 2, we can not find as significant a difference as criterion 1 among three estimators. Every estimator tends to underestimate the dimension in some cases (see Case 1, 7) and overestimate it in other cases (see Case 10).

In most cases in Table 4, the estimation for dimension is made correctly. However, despite a high degree of freedom, we still observe the tendency of $\hat{\tau}^{(0)}$ to underestimation (see Case 1) and that of $\hat{\tau}^{(1)}$ or $\hat{\tau}^{(2)}$ to overestimation (see Case 10) with respect to the criterion 1.

Though both underestimation and overestimation are undesirable, the former is more crucial, and have more substantial effect on the results obtained in principal component analysis or factor analysis, since important component (factor) is neglected (see e.g. the comment in p 278 in Fabriger et. al. (1999)). In this sense, $\hat{\tau}^{(1)}$ and $\hat{\tau}^{(2)}$ are superior to $\hat{\tau}^{(0)}$. The tendency to overestimation of $\hat{\tau}^{(1)}$ is weakened in $\hat{\tau}^{(2)}$ since $\beta_i^{(2)}$'s are closer to unit. We can correct the overestimation further by selecting β_i 's that are much closer to unit, but still satisfy the three conditions in Theorem 2.

4 Conclusion

We can summarize the results of this paper as follows;

1. The distribution of the sample contribution rates is identical within a family of elliptically contoured distributions. It is determined solely by the population contribution rates.
2. A class of new estimators of the population contribution rates was derived. In the estimation of the normalized population covariance matrix, the estimator composed of the new estimator and the sample eigenvectors dominates the estimator composed of the classical estimator and the sample eigenvectors with respect to the entropy loss function.

3. A simulation study shows that the new estimators perform substantially better than the classical estimator with respect to the risk derived from a quadratic loss function. Another simulation study shows that the new estimators tend to overestimate the dimension. They are more suitable than the classical estimator for the decision of dimension in principal component analysis or factor analysis, since the classical estimator is likely to underestimate the dimension.

5 Acknowledgment

We really appreciate many constructive comments and suggestions from two anonymous referees. They broadened the author's vision and improved the quality of this paper.

6 Appendix

6.1 Proof of (5)

We can suppose $\mathbf{S} \sim W_p(n, \mathbf{\Lambda})$, $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$, since the distribution of d_i 's depends only on $\boldsymbol{\lambda}$. Let $\mathbf{\Delta} = (\delta_{ij})$ be defined as

$$\mathbf{\Delta} = n^{-1}\mathbf{S} - \mathbf{\Lambda} = \mathbf{A} - \mathbf{\Lambda}.$$

(3) of Lawley (1959) gives the following expansion of l_r^* ($1 \leq r \leq p$),

$$\begin{aligned} l_r^* &= \lambda_r + \delta_{rr} + \sum_{i \neq r} \frac{\delta_{ri}^2}{\lambda_r - \lambda_i} \\ &\quad - \delta_{rr} \sum_{i \neq r} \left(\frac{\delta_{ri}}{\lambda_r - \lambda_i} \right)^2 + \sum_{i \neq r} \sum_{j \neq r} \left(\frac{\delta_{ri} \delta_{rj} \delta_{ij}}{(\lambda_r - \lambda_i)(\lambda_r - \lambda_j)} \right) + O(\|\boldsymbol{\delta}\|^4). \end{aligned} \quad (41)$$

Since $\sum_{i=1}^p l_i^* = \sum_{i=1}^p (\lambda_i + \delta_{ii})$, we have the Taylor expansion of $(\sum_{i=1}^p l_i^*)^{-1}$

$$\begin{aligned} \frac{1}{\sum_{i=1}^p l_i^*} &= \frac{1}{\sum_{i=1}^p (\lambda_i + \delta_{ii})} \\ &= \frac{1}{\sum_{i=1}^p \lambda_i} - \frac{\sum_{i=1}^p \delta_{ii}}{(\sum_{i=1}^p \lambda_i)^2} + 2 \frac{\sum_{i < j} \delta_{ii} \delta_{jj}}{(\sum_{i=1}^p \lambda_i)^3} \\ &\quad + \frac{\sum_{i=1}^p \delta_{ii}^2}{(\sum_{i=1}^p \lambda_i)^3} - \frac{\sum_{1 \leq i, j, k \leq p} \delta_{ii} \delta_{jj} \delta_{kk}}{(\sum_{i=1}^p \lambda_i)^4} + O(\|\boldsymbol{\delta}\|^4). \end{aligned} \quad (42)$$

Combining (41) and (42), we have

$$\begin{aligned}
d_r &= \frac{l_r^*}{\sum_{i=1}^p l_i^*} \\
&= \frac{\lambda_r}{\sum_{i=1}^p \lambda_i} + \frac{1}{\sum_{i=1}^p \lambda_i} \delta_{rr} - \frac{\lambda_r}{(\sum_{i=1}^p \lambda_i)^2} \sum_{i=1}^p \delta_{ii} \\
&\quad + \frac{\lambda_r}{(\sum_{i=1}^p \lambda_i)^3} \left(\sum_{i=1}^p \delta_{ii}^2 + 2 \sum_{i<j} \delta_{ii} \delta_{jj} \right) \\
&\quad - \frac{1}{(\sum_{i=1}^p \lambda_i)^2} \left(\sum_{i=1}^p \delta_{ii} \right) \delta_{rr} + \frac{1}{\sum_{i=1}^p \lambda_i} \sum_{i \neq r} \frac{\delta_{ri}^2}{\lambda_r - \lambda_i} + O(\|\boldsymbol{\delta}\|^3).
\end{aligned} \tag{43}$$

We can easily calculate the low-dimensional moments of $\boldsymbol{\delta}$. They are given as follows;

$$E[\delta_{ii}] = 0, \quad 1 \leq i \leq p. \tag{44}$$

$$E[\delta_{ii} \delta_{jj}] = \begin{cases} 2n^{-1} \lambda_i^2, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{45}$$

$$E[\delta_{ij}^2] = n^{-1} \lambda_i \lambda_j, \quad \text{if } i \neq j. \tag{46}$$

In addition,

$$E[\|\boldsymbol{\delta}\|^k] = O(n^{-(k-1)}). \tag{47}$$

Substituting (44) – (47) into (43), we have the desired result. \blacksquare

6.2 Proof of (10)

Let the singular value decomposition of \mathbf{Z} denoted by

$$\mathbf{Z} = \mathbf{O} \mathbf{D} \mathbf{H}, \quad \mathbf{D} = \text{diag}(d_1, \dots, d_p), \quad d_1 > \dots > d_p > 0, \quad \mathbf{O} \in V_{p,n}, \quad \mathbf{H} \in O(p),$$

where $O(p)$ is the set of p -dimensional orthogonal matrices. From the Jacobian of this decomposition (see e.g. Theorem 5 of Uhlig (1994)), we have

$$d\mathbf{Z} \propto \prod_{i=1}^p d_i^{n-p} \prod_{i<j} (d_i^2 - d_j^2) \mu_{p,p}(d\mathbf{H}) \mu_{p,n}(d\mathbf{O}) d\mathbf{d},$$

where $\mu_{p,p}$ and $\mu_{p,n}$ are the invariant probability measures respectively on $O(p)$ and $V_{p,n}$. Further by the transformation $d_i \rightarrow t_i = d_i^2$ $i = 1, \dots, p$, we have

$$d\mathbf{Z} \propto \prod_{i=1}^p t_i^{(n-p-1)/2} \prod_{i<j} (t_i - t_j) \mu_{p,p}(d\mathbf{H}) \mu_{p,n}(d\mathbf{O}) dt.$$

Notice that

$$\mathbf{S} = \mathbf{H}' \mathbf{T} \mathbf{H}, \quad \mathbf{T} = \text{diag}(t_1, \dots, t_p), \tag{48}$$

hence,

$$\begin{aligned} & f(\text{tr} \mathbf{Z}' \mathbf{Z} \boldsymbol{\Sigma}^{-1}) |\boldsymbol{\Sigma}|^{-n/2} d\mathbf{Z} \\ & \propto f(\text{tr} \mathbf{S} \boldsymbol{\Sigma}^{-1}) |\mathbf{S}|^{(n-p-1)/2} |\boldsymbol{\Sigma}|^{-n/2} \prod_{i < j} (t_i - t_j) \mu_{p,p}(d\mathbf{H}) \mu_{p,n}(d\mathbf{O}) d\mathbf{t}. \end{aligned}$$

If we integrate the right side of the above equation over $V_{p,n}$, we have the following density function of \mathbf{H} and \mathbf{t} with respect to $\mu_{p,p}(d\mathbf{H})d\mathbf{t}$

$$c_0 f(\text{tr} \mathbf{S} \boldsymbol{\Sigma}^{-1}) |\mathbf{S}|^{(n-p-1)/2} |\boldsymbol{\Sigma}|^{-n/2} \prod_{i < j} (t_i - t_j) \quad (49)$$

with some constant c_0 . Combined with a formula about the spectral decomposition (48) (see e.g. (22) on the p105 of Muirhead(1982))

$$d\mathbf{S} \propto \prod_{i < j} (t_i - t_j) \mu_{pp}(d\mathbf{H}) d\mathbf{t},$$

(49) leads to the following density of \mathbf{S} with respect to Lebesgue measure

$$c_1 f(\text{tr} \mathbf{S} \boldsymbol{\Sigma}^{-1}) |\mathbf{S}|^{(n-p-1)/2} |\boldsymbol{\Sigma}|^{-n/2} \quad (50)$$

with a normalizing constant c_1 . ■

References

- [1] Dey, D. K. and Srinivasan, C. (1988). Estimation of a covariance matrix under Stein's loss. *Ann. Statist.*, **13**, 1581-1591.
- [2] Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychol. Methods*, **4**, 272-299.
- [3] Huang, D-Y. and Tseng, S-T. (1992). A decision procedure for determining the number of components in principal component analysis. *J. Statist. Plann. Inference*, **30**, 63-71.
- [4] Johnson, D. E. and Graybill, F. A. (1972). An analysis of a two-way model with interaction and no replication. *J. Am. Stat. Assoc.*, **67**, 862-868.
- [5] Jolliffe, I. T. (2004). *Principal Component Analysis 2nd.ed.*, Springer, New York
- [6] Karoui, N. E. (2009). Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond. *Ann. Applied Probab.*, **19**, 2362-2405.

- [7] Konishi, S. (1977). Asymptotic expansion for the distribution of a function of latent roots of the covariance matrix. *Ann. Inst. Statist. Math.*, **29**, 389-396.
- [8] Kritchman, S. and Nadler, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems*, **94**, 19-32.
- [9] Krzanowski, W. J. (1979). Some exact percentage points of a statistic useful in analysis of variance and principal component analysis. *Technometrics*, **21**, 261-263.
- [10] Lawley, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128-136.
- [11] Mandel, J. (1972). Principal components, analysis of variance and data structure. *Stat. Neerl.*, **26**, 119-129.
- [12] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- [13] Nadler, B. (2008). Finite sample approximation results for principal component analysis: a matrix perturbation approach. *Ann. Statist.*, **36**, 2791-2817.
- [14] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, **17**, 1617-1642.
- [15] Sugiyama, T. and Tong, H. (1976). On a statistic useful in dimensionality reduction in multivariable linear stochastic system. *Commn. Statist.*, **A5**, 711-721.
- [16] Uhlig, H. (1994). On singular Wishart and singular multivariate beta distributions. *Ann. Statist.*, **22**, 395-405.
- [17] Ulfarsson, M. O. and Solo, V. (2008). Dimension estimation in noisy pca with sure and random matrix theory. *IEEE transactions on signal processing*, **56**, 5804-5816.