Statistica Sinica Preprint No: SS-2016-0063R2

Title	Two-sample tests for high-dimension, strongly spiked eigenvalue models
Manuscript ID	SS-2016-0063R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0063
Complete List of Authors	Makoto Aoshima and Kazuyoshi Yata
Corresponding Author	Makoto Aoshima
E-mail	aoshima@math.tsukuba.ac.jp
Notice: Accepted version subject to English editing.	

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

TWO-SAMPLE TESTS FOR HIGH-DIMENSION, STRONGLY SPIKED EIGENVALUE MODELS

Makoto Aoshima and Kazuyoshi Yata

University of Tsukuba

Abstract: We consider two-sample tests for high-dimensional data under two disjoint models: the strongly spiked eigenvalue (SSE) model and the non-SSE (NSSE) model. We provide a general test statistic as a function of a positive-semidefinite matrix. We give sufficient conditions for the test statistic to satisfy a consistency property and to be asymptotically normal. We discuss an optimality of the test statistic under the NSSE model. We also investigate the test statistic under the SSE model by considering strongly spiked eigenstructures and create a new effective test procedure for the SSE model. Finally, we discuss the performance of the classifiers numerically.

Key words and phrases: Asymptotic normality, eigenstructure estimation, large p small n, noise reduction methodology, spiked model.

1. Introduction

A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively low. This is the so-called "HDLSS" or "large p, small n" data, where p is the data dimension, n is the sample size and $p/n \to \infty$. Statistical inference on this type of data is becoming increasingly relevant, especially in the areas of medical diagnostics, engineering, and other big data. Suppose we have independent samples of p-variate random variables from populations $\pi_i,\ i = 1, 2$, with unknown mean vectors $\boldsymbol{\mu}_i$ and unknown positive-definite covariance matrices $\boldsymbol{\Sigma}_i$. We do not assume the normality of the population distributions. The eigen-decomposition of $\boldsymbol{\Sigma}_i$ $(i = 1, 2)$ is given by $\boldsymbol{\Sigma}_i = \boldsymbol{H}_i\boldsymbol{\Lambda}_i\boldsymbol{H}_i^T = \sum_{j=1}^{p} \lambda_{ij}\boldsymbol{h}_{ij}\boldsymbol{h}_{ij}^T$, where $\boldsymbol{\Lambda}_i = \mathrm{diag}(\lambda_{i1}, ..., \lambda_{ip})$ is a diagonal matrix of eigenvalues, $\lambda_{i1} \geq \cdots \geq \lambda_{ip} > 0$, and $\boldsymbol{H}_i = [\boldsymbol{h}_{i1}, ..., \boldsymbol{h}_{ip}]$ is an orthogonal matrix of the corresponding eigenvectors. Note that λ_{i1} is the largest eigenvalue of $\boldsymbol{\Sigma}_i$ for $i = 1, 2$. For the eigenvalues, we consider two disjoint models: the strongly spiked eigenvalue (SSE) model, which will be defined by (1.6), and the non-SSE (NSSE) model, which will be defined by (1.4).

In this paper, we consider the two-sample test:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{1.1}$$

Having recorded i.i.d. samples, $\boldsymbol{x}_{ij},\ j = 1, ..., n_i$, of size n_i from each π_i, we define $\overline{\boldsymbol{x}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$ and $\boldsymbol{S}_{in_i} = \sum_{j=1}^{n_i}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{in_i})(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{in_i})^T/(n_i - 1)$ for $i = 1, 2$. We assume $n_i \geq 4$ for $i = 1, 2$. Hotelling's T^2-statistic is

$$T^2 = (n_1 + n_2)^{-1}n_1n_2(\overline{\boldsymbol{x}}_{1n_1} - \overline{\boldsymbol{x}}_{2n_2})^T\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_{1n_1} - \overline{\boldsymbol{x}}_{2n_2}),$$

where $\boldsymbol{S} = \{(n_1 - 1)\boldsymbol{S}_{1n_1} + (n_2 - 1)\boldsymbol{S}_{2n_2}\}/(n_1 + n_2 - 2)$. However, $\boldsymbol{S}^{-1}$ does not exist in such HDLSS contexts as $p/n_i \to \infty,\ i = 1, 2$. In such

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

situations, Dempster (1958, 1960) and Srivastava (2007) considered the test when π_1 and π_2 are Gaussian. When π_1 and π_2 are non-Gaussian, Bai and Saranadasa (1996) and Cai, Liu, and Xia (2014) considered the test under homoscedasticity, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Chen and Qin (2010) and Aoshima and Yata (2011, 2015) considered the test under heteroscedasticity, $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$.

In this paper, we first consider a test statistic with a positive-semidefinite matrix $\boldsymbol{A}$ of dimension p:

$$\begin{aligned} T(\boldsymbol{A}) &= (\overline{\boldsymbol{x}}_{1n_1} - \overline{\boldsymbol{x}}_{2n_2})^T \boldsymbol{A} (\overline{\boldsymbol{x}}_{1n_1} - \overline{\boldsymbol{x}}_{2n_2}) - \sum_{i=1}^{2} \mathrm{tr}(\boldsymbol{S}_{in_i}\boldsymbol{A})/n_i \\ &= 2\sum_{i=1}^{2} \frac{\sum_{j<j'}^{n_i} \boldsymbol{x}_{ij}^T \boldsymbol{A} \boldsymbol{x}_{ij'}}{n_i(n_i-1)} - 2\overline{\boldsymbol{x}}_{1n_1}^T \boldsymbol{A} \overline{\boldsymbol{x}}_{2n_2}. \end{aligned} \tag{1.2}$$

Note that $E\{T(\boldsymbol{A})\} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{A} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Let $\boldsymbol{I}_p$ denote the identity matrix of dimension p. We note that $T(\boldsymbol{I}_p)$ is equivalent to the statistics given by Chen and Qin (2010) and Aoshima and Yata (2011). We call the test with $T(\boldsymbol{I}_p)$ the "distance-based two-sample test". In Section 3, we discuss a choice of $\boldsymbol{A}$. We consider the divergence condition $p \to \infty$, $n_1 \to \infty$ and $n_2 \to \infty$, that is equivalent to

$$m \to \infty, \quad \text{where } \ m = \min\{p, n_{\min}\} \ \text{ with } \ n_{\min} = \min\{n_1, n_2\}.$$

By using Theorem 1 in Chen and Qin (2010), or Theorem 4 in Aoshima and Yata (2015), we can claim that under H_0 in (1.1),

$$T(\boldsymbol{I}_p)/\{K_1(\boldsymbol{I}_p)\}^{1/2} \Rightarrow N(0,1) \ \text{ as } m \to \infty \tag{1.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

if we assume (A-i), see Section 2, and the condition that

$$\frac{\lambda_{i1}^2}{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)} \to 0 \quad \text{as } p \to \infty \text{ for } i = 1, 2. \tag{1.4}$$

Here, $K_1(\boldsymbol{A})$ is defined in Section 2.1, "$\Rightarrow$" denotes convergence in distribution and $N(0,1)$ denotes the standard normal. Thus, by using $T(\boldsymbol{I}_p)$ and an estimate of $K_1(\boldsymbol{I}_p)$, one can construct a test procedure of (1.1) for high-dimensional data. As discussed in Section 2 of Aoshima and Yata (2015), the distance-based two-sample test is quite flexible for high-dimension, non-Gaussian data. In Section 3, we investigate an optimality of the test statistic in (1.2) and discuss a choice of $\boldsymbol{A}$.

Remark 1. If all λ_{ij}'s are bounded as $\limsup_{p\to\infty} \lambda_{ij} < \infty$ and $\liminf_{p\to\infty} \lambda_{ij} > 0$, (1.4) trivially holds. On the other hand, they often have a spiked model such as

$$\lambda_{ij} = a_{ij} p^{\alpha_{ij}} \ (j = 1, ..., t_i) \ \text{ and } \ \lambda_{ij} = c_{ij} \ (j = t_i + 1, ..., p), \tag{1.5}$$

where the a_{ij}'s, c_{ij}'s and α_{ij}'s are positive fixed constants and the t_i's are positive fixed integers. If they satisfy (1.5), (1.4) holds when $\alpha_{i1} < 1/2$ for $i = 1, 2$. See Yata and Aoshima (2012) for the details.

For eigenvalues of high-dimensional data, Jung and Marron (2009), Yata and Aoshima (2012, 2013b), Onatski (2012), and Fan, Liao, and Mincheva (2013) considered spiked models such that $\lambda_{ij} \to \infty$ as $p \to \infty$ for $j = 1, ..., k_i$, with some positive integer k_i. The above references show that spiked models are quite natural because the first several eigenvalues

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

should be spiked for high-dimensional data. Hence, we consider the following situation as well:

$$\liminf_{p\to\infty}\left\{\frac{\lambda_{i1}^2}{\operatorname{tr}(\boldsymbol{\Sigma}_i^2)}\right\}>0 \ \text{ for } i=1 \text{ or } 2. \tag{1.6}$$

In (1.6), the first eigenvalue is more spiked than in (1.4). For example, (1.6) holds for the spiked model in (1.5) with $\alpha_{i1}\geq 1/2$. We call (1.6) the "strongly spiked eigenvalue (SSE) model". We emphasize that the asymptotic normality in (1.3) is not satisfied under the SSE model. See Section 4.1. See also Katayama, Kano, and Srivastava (2013) and Ma, Lan, and Wang (2015). Recall that (1.3) holds under (1.4). We call (1.4) the "non-strongly spiked eigenvalue (NSSE) model".

The organization of this paper is as follows. In Section 2, we give sufficient conditions for $T(\boldsymbol{A})$ to satisfy a consistency property and asymptotic normality. In Section 3, under the NSSE model, we give a test procedure with $T(\boldsymbol{A})$ and discuss the choice of $\boldsymbol{A}$. In Section 4, under the SSE model, we investigate test procedures by considering strongly spiked eigenstructures. In Section 5, we create a new test procedure by estimating the eigenstructures for the SSE model. We show that the power of the new test procedure is much higher than the distance-based two-sample test for the SSE model. In Section 6, we discuss the performance of the test procedures for the SSE model with simulations. In Section 7, we highlight the benefits of the new models. In the online supplementary material, we give additional simulations, data analyses, and proofs of the theoretical results. We also provide a method to distinguish between the NSSE model and the

SSE model, and estimate the required parameters.

2. Asymptotic Properties of $T(\boldsymbol{A})$

In this section, we give sufficient conditions for $T(\boldsymbol{A})$ to satisfy a consistency property and to be asymptotically normal. For a positive-semidefinite matrix $\boldsymbol{A}$, we write the square root of $\boldsymbol{A}$ as $\boldsymbol{A}^{1/2}$. Let $\boldsymbol{x}_{ij} = \boldsymbol{H}_i \boldsymbol{\Lambda}_i^{1/2} \boldsymbol{z}_{ij} + \boldsymbol{\mu}_i$, where $\boldsymbol{z}_{ij} = (z_{i1j}, ..., z_{ipj})^T$ is considered as a sphered data vector having the zero mean vector and identity covariance matrix. We assume that the fourth moments of each variable in $\boldsymbol{z}_{ij}$ are uniformly bounded. More specifically, we assume that

$$\boldsymbol{x}_{ij} = \boldsymbol{\Gamma}_i \boldsymbol{w}_{ij} + \boldsymbol{\mu}_i \ \text{ for } i = 1, 2; \ j = 1, ..., n_i, \tag{2.1}$$

where $\boldsymbol{\Gamma}_i$ is a $p \times r_i$ matrix for some $r_i \geq p$ such that $\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i^T = \boldsymbol{\Sigma}_i$, and $\boldsymbol{w}_{ij}$, $j = 1, ..., n_i$, are i.i.d. random vectors having $E(\boldsymbol{w}_{ij}) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{w}_{ij}) = \boldsymbol{I}_{r_i}$. Note that (2.1) includes the case that $\boldsymbol{\Gamma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i^{1/2}$ and $\boldsymbol{w}_{ij} = \boldsymbol{z}_{ij}$. Refer to Bai and Saranadasa (1996), Chen and Qin (2010) and Aoshima and Yata (2015) for the details of the model. As for $\boldsymbol{w}_{ij} = (w_{i1j}, ..., w_{ir_ij})^T$, we assume the following assumption for π_i, $i = 1, 2$, as necessary.

(A-i) The fourth moments of each variable in $\boldsymbol{w}_{ij}$ are uniformly bounded, $E(w_{isj}^2 w_{itj}^2) = E(w_{isj}^2) E(w_{itj}^2)$ and $E(w_{isj} w_{itj} w_{iuj} w_{ivj}) = 0$ for all $s \neq t, u, v$.

When the π_is are Gaussian, (A-i) naturally holds.

2.1. Consistency and Asymptotic Normality of $T(\boldsymbol{A})$

Let $\boldsymbol{\mu}_A = \boldsymbol{A}^{1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, $\boldsymbol{\Sigma}_{i,A} = \boldsymbol{A}^{1/2}\boldsymbol{\Sigma}_i\boldsymbol{A}^{1/2},\ i = 1, 2$, and $\Delta(\boldsymbol{A}) = ||\boldsymbol{\mu}_A||^2$, where $||\cdot||$ denotes the Euclidean norm. Let $K(\boldsymbol{A}) = K_1(\boldsymbol{A}) + K_2(\boldsymbol{A})$, where

$$K_1(\boldsymbol{A}) = 2\sum_{i=1}^{2} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{i,A}^2)}{n_i(n_i-1)} + 4\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{1,A}\boldsymbol{\Sigma}_{2,A})}{n_1 n_2} \ \text{ and } \ K_2(\boldsymbol{A}) = 4\sum_{i=1}^{2} \frac{\boldsymbol{\mu}_A^T\boldsymbol{\Sigma}_{i,A}\boldsymbol{\mu}_A}{n_i}.$$

Here $E\{T(\boldsymbol{A})\} = \Delta(\boldsymbol{A})$ and $\mathrm{Var}\{T(\boldsymbol{A})\} = K(\boldsymbol{A})$. Also, $\Delta(\boldsymbol{A}) = 0$ under H_0 in (1.1). Let $\lambda_{\max}(\boldsymbol{B})$ denote the largest eigenvalue of a positive-semidefinite matrix, $\boldsymbol{B}$. We assume the following condition of the $\boldsymbol{\Sigma}_{i,A}$'s, as necessary.

(A-ii) $\dfrac{\{\lambda_{\max}(\boldsymbol{\Sigma}_{i,A})\}^2}{\mathrm{tr}(\boldsymbol{\Sigma}_{i,A}^2)} \to 0$ as $p \to \infty$ for $i = 1, 2$.

When $\boldsymbol{A} = \boldsymbol{I}_p$, (A-ii) is (1.4). We assume one of the following conditions, as necessary.

(A-iii) $\dfrac{K_1(\boldsymbol{A})}{\{\Delta(\boldsymbol{A})\}^2} \to 0$ as $m \to \infty$; **(A-iv)** $\limsup_{m\to\infty} \dfrac{\{\Delta(\boldsymbol{A})\}^2}{K_1(\boldsymbol{A})} < \infty$;

(A-v) $\dfrac{K_1(\boldsymbol{A})}{K_2(\boldsymbol{A})} \to 0$ as $m \to \infty$.

Note that (A-iv) holds under H_0 in (1.1). If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ ($= \boldsymbol{\Sigma}$, say), (A-iii) holds when $\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}/\{n_{\min}\Delta(\boldsymbol{A})\}^2 \to 0$ as $m \to \infty$. On the other hand, (A-iv) holds when $\liminf_{m\to\infty} \mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}/\{n_{\min}\Delta(\boldsymbol{A})\}^2 > 0$. See Section 3.2 for the details of (A-v).

Proposition 1. *(A-v) implies (A-iii).*

Theorem 1. *If (A-iii) holds, then* $T(\boldsymbol{A})/\Delta(\boldsymbol{A}) = 1 + o_P(1)$ *as* $m \to \infty$.

Theorem 2. *If (A-i) and either (A-ii) and (A-iv) or (A-v) hold, then* $\{T(\boldsymbol{A}) - \Delta(\boldsymbol{A})\}/\{K(\boldsymbol{A})\}^{1/2} \Rightarrow N(0,1)$ *as* $m \to \infty$.

Lemma 1. *If (A-ii) and (A-iv) hold, then* $K(\boldsymbol{A})/K_1(\boldsymbol{A}) = 1 + o(1)$ *as* $m \to \infty$.

Since the $\boldsymbol{\Sigma}_i$'s are unknown, it is necessary to estimate $K_1(\boldsymbol{A})$. Consider the estimator

$$\widehat{K}_1(\boldsymbol{A}) = 2\sum_{i=1}^{2} \frac{W_{in_i}(\boldsymbol{A})}{n_i(n_i-1)} + 4\frac{\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{A}\boldsymbol{S}_{2n_2}\boldsymbol{A})}{n_1 n_2},$$

where $W_{in_i}(\boldsymbol{A})$ is defined by (2.2) in Section 2.2.

Lemma 2. *If (A-i) holds, then* $\widehat{K}_1(\boldsymbol{A})/K_1(\boldsymbol{A}) = 1 + o_P(1)$ *as* $m \to \infty$.

By combining Theorem 2 with Lemmas 1 and 2, we have the following result.

Corollary 1. *If (A-i), (A-ii), and (A-iv) hold, then* $\{T(\boldsymbol{A}) - \Delta(\boldsymbol{A})\}/\{\widehat{K}_1(\boldsymbol{A})\}^{1/2} \Rightarrow N(0,1)$ *as* $m \to \infty$.

2.2. Estimation of tr($\boldsymbol{\Sigma}_A^2$)

Throughout this section, we omit the population subscript. Chen, Zhang, and Zhong (2010) considered an unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$, $W_n = \sum_{i\neq j}^{n}(\boldsymbol{x}_i^T\boldsymbol{x}_j)^2/{}_nP_2 - 2\sum_{i\neq j\neq s}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{x}_s/{}_nP_3 + \sum_{i\neq j\neq s\neq t}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_j\boldsymbol{x}_s^T\boldsymbol{x}_t/{}_nP_4$, where ${}_nP_r = n!/(n-r)!$. Aoshima and Yata (2011) and Yata and Aoshima (2013a) gave a different unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$. From these backgrounds, we

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

construct an unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}_A^2)$ as

$$W_n(\boldsymbol{A}) = \sum_{i\neq j}^{n} \frac{(\boldsymbol{x}_i^T \boldsymbol{A}\boldsymbol{x}_j)^2}{{}_nP_2} - 2\sum_{i\neq j\neq s}^{n} \frac{\boldsymbol{x}_i^T \boldsymbol{A}\boldsymbol{x}_j \boldsymbol{x}_j^T \boldsymbol{A}\boldsymbol{x}_s}{{}_nP_3} + \sum_{i\neq j\neq s\neq t}^{n} \frac{\boldsymbol{x}_i^T \boldsymbol{A}\boldsymbol{x}_j \boldsymbol{x}_s^T \boldsymbol{A}\boldsymbol{x}_t}{{}_nP_4}. \tag{2.2}$$

Note that $E\{W_n(\boldsymbol{A})\} = \mathrm{tr}(\boldsymbol{\Sigma}_A^2)$ and $W_n(\boldsymbol{I}_p) = W_n$. In view of Chen, Zhang, and Zhong (2010), one can claim that

$$\mathrm{Var}\{W_n(\boldsymbol{A})/\mathrm{tr}(\boldsymbol{\Sigma}_A^2)\} \to 0 \tag{2.3}$$

as $p \to \infty$ and $n \to \infty$ under (A-i), so that $W_n(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{\Sigma}_A^2)\{1 + o_P(1)\}$.

3. Test Procedures for Non-Strongly Spiked Eigenvalue Model

In this section, we consider test procedures given by $T(\boldsymbol{A})$ when (A-ii) is met as in the NSSE model. With the help of asymptotic normality, we discuss an optimality of $T(\boldsymbol{A})$ for high-dimensional data.

3.1. Test Procedure by $T(\boldsymbol{A})$

Let z_c be a constant such that $P\{N(0,1) > z_c\} = c$ for $c \in (0,1)$. For given $\alpha \in (0, 1/2)$, from Corollary 1, we consider testing the hypothesis at (1.1) by

$$\text{rejecting } H_0 \Longleftrightarrow T(\boldsymbol{A})/\{\widehat{K}_1(\boldsymbol{A})\}^{1/2} > z_\alpha. \tag{3.1}$$

The power of the test (3.1) depends on $\Delta(\boldsymbol{A})$; we denote it by power$(\Delta(\boldsymbol{A}))$.

Theorem 3. *If (A-i) and (A-ii) hold, then the test (3.1) has, as $m \to \infty$,*

$$\textit{size} = \alpha + o(1) \ \ \textit{and} \ \ \textit{power}(\Delta(\boldsymbol{A})) - \Phi\Big(\frac{\Delta(\boldsymbol{A})}{\{K(\boldsymbol{A})\}^{1/2}} - z_\alpha \Big(\frac{K_1(\boldsymbol{A})}{K(\boldsymbol{A})} \Big)^{1/2} \Big) = o(1),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (c.d.f.) of $N(0,1)$.

Corollary 2. *If (A-i) holds, then, under H_1, the test (3.1) has as, $m \to \infty$,*

$$\begin{aligned} &\textit{power}(\Delta(\boldsymbol{A})) = 1 + o(1) \quad \textit{under (A-iii)}; \\ &\textit{power}(\Delta(\boldsymbol{A})) - \Phi\Big(\frac{\Delta(\boldsymbol{A})}{\{K_1(\boldsymbol{A})\}^{1/2}} - z_\alpha \Big) = o(1) \quad \textit{under (A-ii) and (A-iv)}; \\ &\textit{power}(\Delta(\boldsymbol{A})) - \Phi\Big(\frac{\Delta(\boldsymbol{A})}{\{K_2(\boldsymbol{A})\}^{1/2}} \Big) = o(1) \quad \textit{under (A-v)}. \end{aligned}$$

3.2. Choice of $\boldsymbol{A}$ in (3.1)

Consider the case when (A-v) is met under H_1. From Corollary 2,

$$\text{power}(\Delta(\boldsymbol{A})) \approx \Phi\big(\Delta(\boldsymbol{A})/\{K_2(\boldsymbol{A})\}^{1/2}\big).$$

Let $\boldsymbol{A}_\star = c_\star(\boldsymbol{\Sigma}_1/n_1 + \boldsymbol{\Sigma}_2/n_2)^{-1}$ with $c_\star = 1/n_1 + 1/n_2$. Note that $\boldsymbol{A}_\star = \boldsymbol{\Sigma}^{-1}$ when $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ $(= \boldsymbol{\Sigma})$. Also, note that $\Delta(\boldsymbol{A}_\star) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ $(= \Delta_{MD}$, say) when $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, where $\Delta_{MD}^{1/2}$ is the Mahalanobis distance. Then, from Proposition S1.1 of the supplementary material, $\boldsymbol{A}_\star$ maximizes $\Delta(\boldsymbol{A})/\{K_2(\boldsymbol{A})\}^{1/2}$ over the set of positive-definite matrices of dimension p. Here, consider (A-v). Note that $c_\star^2 p = c_\star^2 \mathrm{tr}\{(\boldsymbol{A}_\star \boldsymbol{A}_\star^{-1})^2\} = \sum_{i=1}^2 \mathrm{tr}\{(\boldsymbol{\Sigma}_i \boldsymbol{A}_\star)^2\}/n_i^2 + 2\mathrm{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{A}_\star \boldsymbol{\Sigma}_2 \boldsymbol{A}_\star)/(n_1 n_2)$, so that $K_1(\boldsymbol{A}_\star) = 2c_\star^2 p\{1 + o(1)\}$ as $m \to \infty$. Also, note that $K_2(\boldsymbol{A}_\star) = 4c_\star \Delta(\boldsymbol{A}_\star)$. Thus, if (A-v)

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

holds,

$$K_1(\boldsymbol{A}_\star)/K_2(\boldsymbol{A}_\star) = O\big(pc_\star/\Delta(\boldsymbol{A}_\star)\big) = O\big(p/\{n_{\min}\Delta(\boldsymbol{A}_\star)\}\big) \to 0 \quad \text{as } m \to \infty.$$

This is severe for high-dimensional data. For example, when $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ and the Mahalanobis distance is bounded as $\limsup_{p\to\infty} \Delta_{MD} < \infty$, the sample size should be large enough that $n_{\min}/p \to \infty$ because $\Delta(\boldsymbol{A}_\star) = \Delta_{MD}$. Hence, (A-v) is quite strict for high-dimensional data. From Proposition 1 and Corollary 2, for any choice of $\boldsymbol{A}$ in (3.1), power$(\Delta(\boldsymbol{A})) = 1+o(1)$ under (A-v). Hence, the optimal choice of $\boldsymbol{A}$ does not make much improvement in the power if (A-v) is met. If (A-v) is not met (i.e., (A-iv) is met), the test (3.1) has

$$\text{power}(\Delta(\boldsymbol{A})) \approx \Phi\big(\Delta(\boldsymbol{A})/\{K_1(\boldsymbol{A})\}^{1/2} - z_\alpha\big)$$

from Corollary 2. In this case, $\boldsymbol{A}_\star$ is not the optimal choice any longer. Because of these reasons, we do not recommend using a test procedure based on the Mahalanobis distance, such as (3.1) with $\boldsymbol{A} = \boldsymbol{A}_\star$. In addition, it is difficult to estimate $\boldsymbol{A}_\star$ for high-dimensional data unless the $\boldsymbol{\Sigma}_i$'s are sparse. When they are sparse, see Bickel and Levina (2008).

Srivastava, Katayama, and Kano (2013) considered a two-sample test using $\boldsymbol{A}_{\star(d)} = c_\star(\boldsymbol{\Sigma}_{1(d)}/n_1+\boldsymbol{\Sigma}_{2(d)}/n_2)^{-1}$ for $\boldsymbol{A}$, where $\boldsymbol{\Sigma}_{i(d)} = \text{diag}(\sigma_{i(1)}, ..., \sigma_{i(p)})$ with $\sigma_{i(j)}$ (> 0) the j-th diagonal element of $\boldsymbol{\Sigma}_i$ for $i = 1, 2;\ j = 1, ..., p$. We do not recommend choosing $\boldsymbol{A}_{\star(d)}$ unless (A-v) is met and the $\boldsymbol{\Sigma}_i$'s are diagonal matrices. If (A-ii) holds, as in the NSSE model, we rather rec-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

ommend choosing $\boldsymbol{A} = \boldsymbol{I}_p$ in (3.1), yielding the distance-based two-sample test. When $\boldsymbol{A} = \boldsymbol{I}_p$, it is not necessary to estimate $\boldsymbol{A}$ and it is quite flexible for high-dimensional, non-Gaussian data. See Section 2 of Aoshima and Yata (2015) for the details.

3.3. Simulations

We used computer simulations to study the performance of the test procedure given by (3.1) when $\boldsymbol{A} = \boldsymbol{I}_p$, $\boldsymbol{A} = \boldsymbol{A}_\star$, $\boldsymbol{A} = \boldsymbol{A}_{\star(d)}$ and $\boldsymbol{A} = \widehat{\boldsymbol{A}}_{\star(d)}$. Here, $\widehat{\boldsymbol{A}}_{\star(d)} = c_\star(\boldsymbol{S}_{1n_1(d)}/n_1 + \boldsymbol{S}_{2n_2(d)}/n_2)^{-1}$, where $\boldsymbol{S}_{in_i(d)} = \mathrm{diag}(s_{in_i1}, ..., s_{in_ip})$, $i = 1, 2$, with s_{in_ij} the j-th diagonal element of $\boldsymbol{S}_{in_i}$. Srivastava, Katayama, and Kano (2013) considered a test procedure given by $T(\widehat{\boldsymbol{A}}_{\star(d)})$. We set $\alpha = 0.05$. Independent pseudo-random observations were generated from $\pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2$. We set $p = 2^s$, $s = 4, ..., 10$ and $n_1 = n_2 = \lceil p^{1/2} \rceil$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. We set $\boldsymbol{\mu}_1 = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{C}(0.3^{|i-j|^{1/2}})\boldsymbol{C}$, where $\boldsymbol{C} = \mathrm{diag}[\{0.5 + 1/(p+1)\}^{1/2}, ..., \{0.5 + p/(p+1)\}^{1/2}]$. We considered three cases: (a) $\boldsymbol{\mu}_2 = \boldsymbol{0}$, (b) $\boldsymbol{\mu}_2 = (1, ..., 1, 0, ..., 0)^T$ whose first ten elements are 1, and (c) $\boldsymbol{\mu}_2 = (0, ..., 0, 1, ..., 1)^T$ whose last ten elements are 1. When $\boldsymbol{A} = \boldsymbol{I}_p$, $\boldsymbol{A} = \boldsymbol{A}_\star$, and $\boldsymbol{A} = \boldsymbol{A}_{\star(d)}$, we note that (A-ii) and (A-iv) are met for (a), (b), and (c).

We checked the performance of the test procedures given by (3.1) with (I) $\boldsymbol{A} = \boldsymbol{I}_p$, (II) $\boldsymbol{A} = \boldsymbol{A}_\star$, (III) $\boldsymbol{A} = \boldsymbol{A}_{\star(d)}$, and (IV) $\boldsymbol{A} = \widehat{\boldsymbol{A}}_{\star(d)}$. The findings were obtained by averaging the outcomes from 2000 ($= R$, say) replications in each situation. We defined $P_r = 1$ (or 0) when H_0 was falsely rejected (or not) for $r = 1, ..., 2000$ for (a) and defined $\overline{\alpha} = \sum_{r=1}^R P_r/R$ to

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

estimate the size. We also defined $P_r = 1$ (or 0) when H_1 was falsely rejected (or not) for $r = 1, ..., 2000$ for (b) and (c) and defined $1 - \overline{\beta} = 1 - \sum_{r=1}^{R} P_r/R$ to estimate the power. Note that their standard deviations are less than 0.011. In Fig. 1, we plotted $\overline{\alpha}$ for (a) and $1 - \overline{\beta}$ for (b) and (c). We also plotted the asymptotic power, $\Phi(\Delta(\boldsymbol{A})/\{K(\boldsymbol{A})\}^{1/2} - z_{\alpha}\{K_1(\boldsymbol{A})/K(\boldsymbol{A})\}^{1/2})$, for (I) to (III) by using Theorem 3. As expected, we

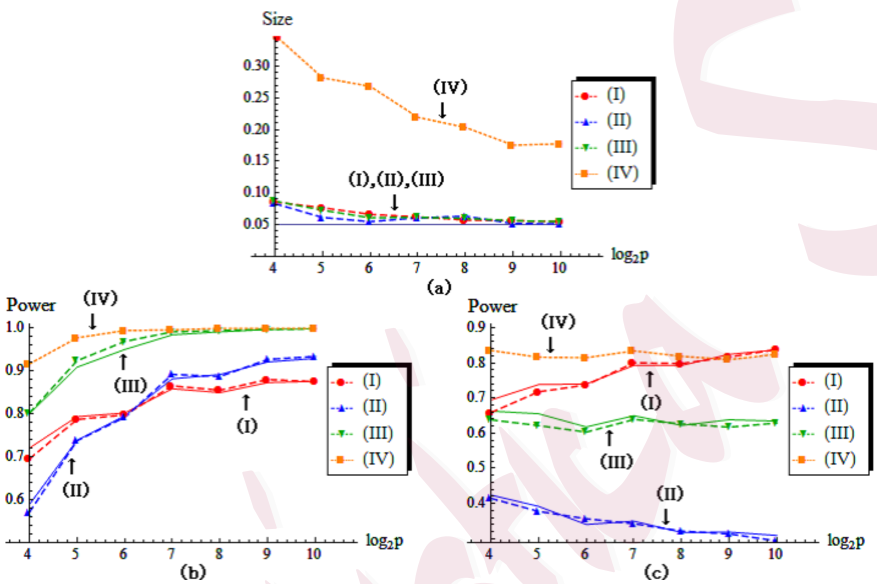


Figure 1: Tests by (3.1) when (I) $\boldsymbol{A} = \boldsymbol{I}_p$, (II) $\boldsymbol{A} = \boldsymbol{A}_{\star}$, (III) $\boldsymbol{A} = \boldsymbol{A}_{\star(d)}$ and (IV) $\boldsymbol{A} = \widehat{\boldsymbol{A}}_{\star(d)}$. The values of $\overline{\alpha}$ are denoted by the dashed lines in the top panel. The values of $1-\overline{\beta}$ are denoted by the dashed lines in the left panel for (b) and in the right panel for (c). The asymptotic powers were given by $\Phi(\Delta(\boldsymbol{A})/\{K(\boldsymbol{A})\}^{1/2} - z_{\alpha}\{K_1(\boldsymbol{A})/K(\boldsymbol{A})\}^{1/2})$ for (I) to (III) which are denoted by the solid lines both in the panels.

observe that the plots get close to the theoretical values. The test with (II) gave a better performance compared to (I) for (b); however, it gave quite a poor performance for (c). The test procedure based on the Mahalanobis distance does not always give a preferable performance for high-dimensional data even when the population distributions are Gaussian with a known

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

and common covariance matrix. See Section 3.2 for the details. We observe that the test with (III) gives a good performance compared to (I) for (b); however, they trade places under (c), because $\Delta(\boldsymbol{I}_p) < \Delta(\boldsymbol{A}_{\star(d)})$ for (b) and $\Delta(\boldsymbol{I}_p) > \Delta(\boldsymbol{A}_{\star(d)})$ for (c) when p is sufficiently large. The test with (IV) gave quite a poor performance because the size for (IV) was much higher than α even when p and the n_i's are large. Hence, we do not recommend using the test procedures based on the Mahalanobis distance or the diagonal matrices unless the n_i's are large enough to claim (A-v).

We also checked the performance of the test procedures by (3.1) for the multivariate skew normal (MSN) distribution. See Azzalini and Dalla Valle (1996) for the details of the MSN distribution. We observed performance similar to that in Fig 1. The results are in Section S4.1 of the supplementary material.

4. Test Procedures for Strongly Spiked Eigenvalue Model

In this section, we consider test procedures when (A-ii) is not met, as in the SSE model. We emphasize that high-dimensional data often obey the SSE model. See Fig. 1 in Yata and Aoshima (2013b) or Section S3 of the supplementary material as well. In case of (A-iv), $T(\boldsymbol{A})$ does not satisfy the asymptotic normality in Theorem 2, so that one cannot use the test (3.1). For example, as for $T(\boldsymbol{I}_p)$, we cannot claim either (1.3) or "size= $\alpha + o(1)$" under the SSE model. In such situations, we consider alternative test procedures.

4.1. Distance-Based Two-Sample Test

We write $T_I = T(\boldsymbol{I}_p)$, $K_{1(I)} = K_1(\boldsymbol{I}_p)$, and $\widehat{K}_{1(I)} = \widehat{K}_1(\boldsymbol{I}_p)$ when $\boldsymbol{A} = \boldsymbol{I}_p$. For the SSE model, Katayama, Kano, and Srivastava (2013) considered a one-sample test. Ma, Lan, and Wang (2015) considered a two-sample test for a factor model which is a special case of the SSE model. Katayama, Kano, and Srivastava (2013) showed that a test statistic is asymptotically distributed as a χ^2 distribution under the Gaussian assumption. For the two-sample test in (1.1), we have the following result.

Theorem 4. *Assume*

$$|\boldsymbol{h}_{11}^T \boldsymbol{h}_{21}| = 1 + o(1) \quad \textit{and} \quad \Psi_{i(2)}/\lambda_{i1}^2 \to 0,\ i = 1, 2,\ \textit{as } p \to \infty, \tag{4.1}$$

where

$$\Psi_{i(s)} = \sum_{j=s}^{p} \lambda_{ij}^2 \quad \textit{for } i = 1, 2;\ s = 1, ..., p.$$

Then, $(2/K_{1(I)})^{1/2} T_I + 1 \Rightarrow \chi_1^2$ *as* $m \to \infty$ *under* H_0*, where* χ_ν^2 *denotes a random variable having a* χ^2 *distribution with* ν *degrees of freedom.*

We test (1.1) by

$$\text{rejecting } H_0 \Longleftrightarrow (2/\widehat{K}_{1(I)})^{1/2} T_I + 1 > \chi_1^2(\alpha), \tag{4.2}$$

where $\chi_1^2(\alpha)$ denotes the $(1-\alpha)$th quantile of χ_1^2. Note that $\widehat{K}_{1(I)}/K_{1(I)} = 1 + o_P(1)$ as $m \to \infty$ under (A-i). Then, from Theorem 4, the test (4.2) ensures that size$= \alpha + o(1)$ as $m \to \infty$ under (A-i).

We note that "$|\boldsymbol{h}_{11}^T \boldsymbol{h}_{21}| = 1 + o(1)$ as $p \to \infty$" in (4.1) is not a general

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

condition for high-dimensional data, so that it is necessary to check the condition in data analyses. See Lemma 4.1 in Ishii, Yata, and Aoshima (2016) for checking the condition. When (4.1) is not met, the test (4.2) cannot ensure accuracy.

4.2. Test Statistics Using Eigenstructures

We consider the following model.

(A-vi) For $i = 1, 2$, there exists a positive fixed integer k_i such that $\lambda_{i1}, ..., \lambda_{ik_i}$ are distinct in the sense that $\liminf_{p\to\infty}(\lambda_{ij}/\lambda_{ij'} - 1) > 0$ when $1 \le j < j' \le k_i$, and λ_{ik_i} and λ_{ik_i+1} satisfy

$$\liminf_{p\to\infty} \frac{\lambda_{ik_i}^2}{\Psi_{i(k_i)}} > 0 \quad \text{and} \quad \frac{\lambda_{ik_i+1}^2}{\Psi_{i(k_i+1)}} \to 0 \quad \text{as } p \to \infty.$$

Note that (A-vi) implies (1.6); (A-vi) is one of the SSE models. (A-vi) is also a power spiked model given by Yata and Aoshima (2013b). For the spiked model in (1.5), (A-vi) holds under the conditions that $\alpha_{ik_i} \ge 1/2$, $a_{ij} \ne a_{ij'}$ for $1 \le j < j' \le k_i$ $(< t_i)$, and $\alpha_{ik_i+1} < 1/2$ for $i = 1, 2$. We consider the following test statistic with positive-semidefinite matrices, $\boldsymbol{A}_i$, $i = 1, 2$, of dimension p:

$$T(\boldsymbol{A}_1, \boldsymbol{A}_2) = 2\sum_{i=1}^{2} \frac{\sum_{j<j'}^{n_i} \boldsymbol{x}_{ij}^T \boldsymbol{A}_i \boldsymbol{x}_{ij'}}{n_i(n_i-1)} - 2\overline{\boldsymbol{x}}_{1n_1}^T \boldsymbol{A}_1^{1/2} \boldsymbol{A}_2^{1/2} \overline{\boldsymbol{x}}_{2n_2}.$$

We do not recommend choosing $\boldsymbol{A}_i = \boldsymbol{\Sigma}_i^{-1}$, $i = 1, 2$; see Section S1.2 in the supplementary material for the details. In addition, it is difficult to estimate $\boldsymbol{\Sigma}_i^{-1}$'s for high-dimensional, non-sparse data. Here, we consider

$\boldsymbol{A}_i$'s as

$$\boldsymbol{A}_{i(k_i)} = \boldsymbol{I}_p - \sum_{j=1}^{k_i} \boldsymbol{h}_{ij}\boldsymbol{h}_{ij}^T = \sum_{j=k_i+1}^{p} \boldsymbol{h}_{ij}\boldsymbol{h}_{ij}^T \quad \text{for } i = 1, 2.$$

Note that $\boldsymbol{A}_{i(k_i)} = \boldsymbol{A}_{i(k_i)}^{1/2}$. We write $\boldsymbol{\mu}_* = \boldsymbol{A}_{1(k_1)}\boldsymbol{\mu}_1 - \boldsymbol{A}_{2(k_2)}\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_{i*} = \boldsymbol{A}_{i(k_i)}\boldsymbol{\Sigma}_i\boldsymbol{A}_{i(k_i)} = \sum_{j=k_i+1}^{p} \lambda_{ij}\boldsymbol{h}_{ij}\boldsymbol{h}_{ij}^T$ for $i = 1, 2$. Let $T_* = T(\boldsymbol{A}_{1(k_1)}, \boldsymbol{A}_{2(k_2)})$, $\Delta_* = ||\boldsymbol{\mu}_*||^2$, and $K_* = K_{1*} + K_{2*}$, where

$$K_{1*} = 2\sum_{i=1}^{2} \frac{\text{tr}(\boldsymbol{\Sigma}_{i*}^2)}{n_i(n_i - 1)} + 4\frac{\text{tr}(\boldsymbol{\Sigma}_{1*}\boldsymbol{\Sigma}_{2*})}{n_1 n_2} \quad \text{and} \quad K_{2*} = 4\sum_{i=1}^{2} \frac{\boldsymbol{\mu}_*^T\boldsymbol{\Sigma}_{i*}\boldsymbol{\mu}_*}{n_i}.$$

Note that $E(T_*) = \Delta_*$ and $\text{Var}(T_*) = K_*$. Also, note that $\text{tr}(\boldsymbol{\Sigma}_{i*}^2) = \Psi_{i(k_i+1)}$ and $\lambda_{\max}(\boldsymbol{\Sigma}_{i*}) = \lambda_{k_i+1}$ for $i = 1, 2$, so that

$$\lambda_{\max}^2(\boldsymbol{\Sigma}_{i*})/\text{tr}(\boldsymbol{\Sigma}_{i*}^2) \to 0 \quad \text{as } p \to \infty \text{ for } i = 1, 2, \text{ under (A-vi)}.$$

From Theorem 2, we have the following result.

Corollary 3. *If (A-i) holds and* $\limsup_{m\to\infty} \Delta_*^2/K_{1*} < \infty$*, then, under (A-vi),* $(T_* - \Delta_*)/K_*^{1/2} \Rightarrow N(0, 1)$ *as* $m \to \infty$.

It does not always hold that $\Delta_* = 0$ under H_0 when $\boldsymbol{A}_{1(k_1)} \neq \boldsymbol{A}_{2(k_2)}$. We assume the following.

(A-vii) $\dfrac{\Delta_*^2}{K_{1*}} \to 0$ as $m \to \infty$ under H_0.

This is a mild condition because $\boldsymbol{A}_{1(k_1)} - \boldsymbol{A}_{2(k_2)} = \sum_{j=1}^{k_2} \boldsymbol{h}_{2j}\boldsymbol{h}_{2j}^T - \sum_{j=1}^{k_1} \boldsymbol{h}_{1j}\boldsymbol{h}_{1j}^T$ is a low-rank matrix with rank $k_1 + k_2$ at most, and under H_0 $\Delta_* = ||(\boldsymbol{A}_{1(k_1)} - \boldsymbol{A}_{2(k_2)})\boldsymbol{\mu}_1||^2$ is small. From Corollary 3, under H_0, it follows that $P(T_*/K_{1*}^{1/2} > z_\alpha) = \alpha + o(1)$. Similar to (3.1), one can construct a test procedure by using T_*. Let

$$x_{ijl} = \boldsymbol{h}_{ij}^T \boldsymbol{x}_{il} = \lambda_{ij}^{1/2} z_{ijl} + \mu_{i(j)} \quad \text{for all } i, j, l, \text{ where } \mu_{i(j)} = \boldsymbol{h}_{ij}^T \boldsymbol{\mu}_i.$$

Then, we write that

$$\begin{aligned} T_* =& 2\sum_{i=1}^{2} \frac{\sum_{l<l'}^{n_i} (\boldsymbol{x}_{il}^T \boldsymbol{x}_{il'} - \sum_{j=1}^{k_i} x_{ijl} x_{ijl'})}{n_i(n_i-1)} \\ &- 2\frac{\sum_{l=1}^{n_1}\sum_{l'=1}^{n_2} (\boldsymbol{x}_{1l} - \sum_{j=1}^{k_1} x_{1jl}\boldsymbol{h}_{1j})^T (\boldsymbol{x}_{2l'} - \sum_{j=1}^{k_2} x_{2jl'}\boldsymbol{h}_{2j})}{n_1 n_2}. \end{aligned}$$

In order to use T_*, it is necessary to estimate the x_{ijl}'s and $\boldsymbol{h}_{ij}$'s.

5. Test Procedure Using Eigenstructures for Strongly Spiked Eigenvalue Model

In this section, we assume (A-vi) and the following for the π_i's:

(A-viii) $E(z_{isj}^2 z_{itj}^2) = E(z_{isj}^2)E(z_{itj}^2)$, $E(z_{isj} z_{itj} z_{iuj}) = 0$ and $E(z_{isj} z_{itj} z_{iuj} z_{ivj}) = 0$ for all $s \neq t, u, v$, with z_{ijl}'s defined in Section 2.

Note that (A-viii) implies (A-i) because the $E(z_{ijl}^4)$'s are bounded and (2.1) includes the case that $\boldsymbol{\Gamma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i^{1/2}$ and $\boldsymbol{w}_{ij} = \boldsymbol{z}_{ij}$. When the π_i's are Gaussian, (A-viii) naturally holds.

5.1. Estimation of Eigenvalues and Eigenvectors

Throughout this section, we omit the population subscript for the sake

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

of simplicity. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$ be the eigenvalues of $\boldsymbol{S}_n$, and write the eigen-decomposition of $\boldsymbol{S}_n$ as $\boldsymbol{S}_n = \sum_{j=1}^{p} \hat{\lambda}_j \hat{\boldsymbol{h}}_j \hat{\boldsymbol{h}}_j^T$, where $\hat{\boldsymbol{h}}_j$ denotes a unit eigenvector corresponding to $\hat{\lambda}_j$. We assume $\boldsymbol{h}_j^T \hat{\boldsymbol{h}}_j \geq 0$ w.p.1 for all j, without loss of generality. Let $\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]$ and $\overline{\boldsymbol{X}} = [\overline{\boldsymbol{x}}_n, ..., \overline{\boldsymbol{x}}_n]$. Then, we define the $n \times n$ dual sample covariance matrix by $\boldsymbol{S}_D = (n-1)^{-1}(\boldsymbol{X} - \overline{\boldsymbol{X}})^T(\boldsymbol{X} - \overline{\boldsymbol{X}})$. Note that $\boldsymbol{S}_n$ and $\boldsymbol{S}_D$ share non-zero eigenvalues. We write the eigen-decomposition of $\boldsymbol{S}_D$ as $\boldsymbol{S}_D = \sum_{j=1}^{n-1} \hat{\lambda}_j \hat{\boldsymbol{u}}_j \hat{\boldsymbol{u}}_j^T$, where $\hat{\boldsymbol{u}}_j = (\hat{u}_{j1}, ..., \hat{u}_{jn})^T$ denotes a unit eigenvector corresponding to $\hat{\lambda}_j$. Note that $\hat{\boldsymbol{h}}_j$ can be calculated as $\hat{\boldsymbol{h}}_j = \{(n-1)\hat{\lambda}_j\}^{-1/2}(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_j$. Let $\delta_j = \lambda_j^{-1} \sum_{s=k+1}^{p} \lambda_s/(n-1)$, for $j = 1, ..., k$. Let $m_0 = \min\{p, n\}$.

Proposition 2. *If (A-vi) and (A-viii) hold, then for $j = 1, ..., k$, $\hat{\lambda}_j/\lambda_j = 1 + \delta_j + O_P(n^{-1/2})$ and $(\hat{\boldsymbol{h}}_j^T \boldsymbol{h}_j)^2 = (1 + \delta_j)^{-1} + O_P(n^{-1/2})$ as $m_0 \to \infty$.*

If $\delta_j \to \infty$ as $m_0 \to \infty$, $\hat{\lambda}_j$ and $\hat{\boldsymbol{h}}_j$ are strongly inconsistent in the sense that $\lambda_j/\hat{\lambda}_j = o_P(1)$ and $(\hat{\boldsymbol{h}}_j^T \boldsymbol{h}_j)^2 = o_P(1)$. See Jung and Marron (2009) for the concept of the strong inconsistency. Also, from Proposition 2, under (A-vi) and (A-viii), as $m_0 \to \infty$,

$$||\hat{\boldsymbol{h}}_j - \boldsymbol{h}_j||^2 = 2\{1 - (1 + \delta_j)^{-1/2}\} + O_P(n^{-1/2}) \quad \text{for } j = 1, ..., k. \tag{5.1}$$

In order to overcome the curse of dimensionality, Yata and Aoshima (2012) proposed an eigenvalue estimation called the noise-reduction (NR) methodology, which was brought about by a geometric representation of $\boldsymbol{S}_D$. If one applies the NR methodology, the λ_j's are estimated by

$$\tilde{\lambda}_j = \hat{\lambda}_j - \frac{\mathrm{tr}(\boldsymbol{S}_D) - \sum_{l=1}^{j} \hat{\lambda}_l}{n-1-j} \quad (j = 1, ..., n-2). \tag{5.2}$$

Here $\tilde{\lambda}_j \geq 0$ w.p.1 for $j = 1, ..., n-2$, and the second term in (5.2) is an estimator of $\lambda_j \delta_j$. When applying the NR methodology to the PC direction vector, one obtains

$$\tilde{\boldsymbol{h}}_j = \{(n-1)\tilde{\lambda}_j\}^{-1/2}(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_j \tag{5.3}$$

for $j = 1, ..., n-2$.

Proposition 3. *If (A-vi) and (A-viii) hold, then for* $j = 1, ..., k$, $\tilde{\lambda}_j/\lambda_j = 1 + O_P(n^{-1/2})$ *and* $(\tilde{\boldsymbol{h}}_j^T \boldsymbol{h}_j)^2 = 1 + O_P(n^{-1})$ *as* $m_0 \to \infty$.

Here $\tilde{\boldsymbol{h}}_j$ is not a unit vector because $||\tilde{\boldsymbol{h}}_j||^2 = \hat{\lambda}_j/\tilde{\lambda}_j$. From Propositions 2 and 3, under (A-vi) and (A-viii), $||\tilde{\boldsymbol{h}}_j - \boldsymbol{h}_j||^2 = \delta_j\{1 + o_P(1)\} + O_P(n^{-1/2})$ as $m_0 \to \infty$ for $j = 1, ..., k$. We note that $2\{1 - (1+\delta_j)^{-1/2}\} < \delta_j$. Thus, in view of (5.1), the norm loss of $\tilde{\boldsymbol{h}}_j$ is larger than that of $\hat{\boldsymbol{h}}_j$. However, $\tilde{\boldsymbol{h}}_j$ is a consistent estimator of $\boldsymbol{h}_j$ in terms of the inner product even when $\delta_j \to \infty$ as $m_0 \to \infty$.

We note that $\boldsymbol{h}_j^T(\boldsymbol{x}_l - \boldsymbol{\mu}) = \lambda_j^{1/2} z_{jl}$ for all j, l. For $\hat{\boldsymbol{h}}_j$ and $\tilde{\boldsymbol{h}}_j$, we have the following result.

Proposition 4. *If (A-vi) and (A-viii) hold, then for* $j = 1, ..., k$ $(l = 1, ..., n)$, $\lambda_j^{-1/2}\hat{\boldsymbol{h}}_j^T(\boldsymbol{x}_l - \boldsymbol{\mu}) = (1+\delta_j)^{-1/2}[z_{jl} + (n-1)^{1/2}\hat{u}_{jl}\delta_j\{1 + o_P(1)\}] + O_P(n^{-1/2})$ *and* $\lambda_j^{-1/2}\tilde{\boldsymbol{h}}_j^T(\boldsymbol{x}_l - \boldsymbol{\mu}) = z_{jl} + (n-1)^{1/2}\hat{u}_{jl}\delta_j\{1 + o_P(1)\} + O_P(n^{-1/2})$ *as* $m_0 \to \infty$.

Consider the standard deviation of these quantities. Note that $[\sum_{l=1}^n\{(n-1)^{1/2}\hat{u}_{jl}\delta_j\}^2/n]^{1/2} = O(\delta_j)$ and $\delta_j = O\{p/(n\lambda_j)\}$ for $\lambda_{k+1} = O(1)$. Hence, in Proposition 4, the inner products are quite biased when p is large, as follows. Let $\boldsymbol{P}_n = \boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n^T/n$, where $\boldsymbol{1}_n = (1,...,1)^T$. Then, $\boldsymbol{1}_n^T\hat{\boldsymbol{u}}_j = 0$ and $\boldsymbol{P}_n\hat{\boldsymbol{u}}_j = \hat{\boldsymbol{u}}_j$ when $\hat{\lambda}_j > 0$, since $\boldsymbol{1}_n^T\boldsymbol{S}_D\boldsymbol{1}_n = 0$. Also, when $\hat{\lambda}_j > 0$,

$$\{(n-1)\tilde{\lambda}_j\}^{1/2}\tilde{\boldsymbol{h}}_j = (\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_j = (\boldsymbol{X} - \boldsymbol{M})\boldsymbol{P}_n\hat{\boldsymbol{u}}_j = (\boldsymbol{X} - \boldsymbol{M})\hat{\boldsymbol{u}}_j,$$

where $\boldsymbol{M} = [\boldsymbol{\mu},...,\boldsymbol{\mu}]$. Thus it holds that $\{(n-1)\tilde{\lambda}_j\}^{1/2}\tilde{\boldsymbol{h}}_j^T(\boldsymbol{x}_l - \boldsymbol{\mu}) = \hat{\boldsymbol{u}}_j^T(\boldsymbol{X} - \boldsymbol{M})^T(\boldsymbol{x}_l - \boldsymbol{\mu}) = \hat{u}_{jl}||\boldsymbol{x}_l - \boldsymbol{\mu}||^2 + \sum_{s=1(\neq l)}^n \hat{u}_{js}(\boldsymbol{x}_s - \boldsymbol{\mu})^T(\boldsymbol{x}_l - \boldsymbol{\mu})$, so that $\hat{u}_{jl}||\boldsymbol{x}_l - \boldsymbol{\mu}||^2$ is biased since $E(||\boldsymbol{x}_l - \boldsymbol{\mu}||^2)/\{(n-1)^{1/2}\lambda_j\} \geq (n-1)^{1/2}\delta_j$. Hence, one should not apply the $\hat{\boldsymbol{h}}_j$'s or the $\tilde{\boldsymbol{h}}_j$'s to the estimation of the inner product.

Consider a bias-reduced estimation of the inner product. Write

$$\hat{\boldsymbol{u}}_{jl} = (\hat{u}_{j1},...,\hat{u}_{jl-1},-\hat{u}_{jl}/(n-1),\hat{u}_{jl+1},...,\hat{u}_{jn})^T$$

with l-th element $-\hat{u}_{jl}/(n-1)$ for all j,l. Note that $\hat{\boldsymbol{u}}_{jl} = \hat{\boldsymbol{u}}_j - (0,...,0,\{n/(n-1)\}\hat{u}_{jl},0,...,0)^T$ and $\sum_{l=1}^n \hat{\boldsymbol{u}}_{jl}/n = \{(n-2)/(n-1)\}\hat{\boldsymbol{u}}_j$. Let

$$c_n = (n-1)^{1/2}/(n-2) \quad \text{and} \quad \tilde{\boldsymbol{h}}_{jl} = c_n\tilde{\lambda}_j^{-1/2}(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_{jl} \tag{5.4}$$

for all j,l. Here $\sum_{l=1}^n \tilde{\boldsymbol{h}}_{jl}/n = \tilde{\boldsymbol{h}}_j$. When $\hat{\lambda}_j > 0$, $c_n^{-1}\tilde{\lambda}_j^{1/2}\tilde{\boldsymbol{h}}_{jl} = (\boldsymbol{X} - \boldsymbol{M})\boldsymbol{P}_n\hat{\boldsymbol{u}}_{jl} = (\boldsymbol{X} - \boldsymbol{M})\hat{\boldsymbol{u}}_{j(l)}$ since $\boldsymbol{1}_n^T\hat{\boldsymbol{u}}_j = \sum_{l=1}^n \hat{u}_{jl} = 0$, where

$$\hat{\boldsymbol{u}}_{j(l)} = (\hat{u}_{j1},...,\hat{u}_{jl-1},0,\hat{u}_{jl+1},...,\hat{u}_{jn})^T + (n-1)^{-1}\hat{u}_{jl}\boldsymbol{1}_{n(l)} \quad \text{for } l = 1,...,n.$$

Here, $\mathbf{1}_{n(l)} = (1, ..., 1, 0, 1, ..., 1)^T$ whose l-th element is 0. Thus it holds that

$$\begin{aligned} c_n^{-1}\tilde{\lambda}_j^{1/2}\tilde{\boldsymbol{h}}_{jl}^T(\boldsymbol{x}_l - \boldsymbol{\mu}) &= \hat{\boldsymbol{u}}_{j(l)}^T(\boldsymbol{X} - \boldsymbol{M})^T(\boldsymbol{x}_l - \boldsymbol{\mu}) \\ &= \sum_{s=1(\neq l)}^{n} \{\hat{u}_{js} + (n-1)^{-1}\hat{u}_{jl}\}(\boldsymbol{x}_s - \boldsymbol{\mu})^T(\boldsymbol{x}_l - \boldsymbol{\mu}), \end{aligned}$$

so that the large biased term, $||\boldsymbol{x}_l - \boldsymbol{\mu}||^2$, has vanished.

Proposition 5. *If (A-vi) and (A-viii) hold, then for* $j = 1, ..., k$ $(l = 1, ..., n)$, $\lambda_j^{-1/2}\tilde{\boldsymbol{h}}_{jl}^T(\boldsymbol{x}_l - \boldsymbol{\mu}) = z_{jl} + \hat{u}_{jl} \times O_P\{(n^{1/2}\lambda_j)^{-1}\lambda_1\} + O_P(n^{-1/2})$ *as* $m_0 \to \infty$.

As $[\sum_{l=1}^n \{\hat{u}_{jl}\lambda_1/(n^{1/2}\lambda_j)\}^2/n]^{1/2} = \lambda_1/(\lambda_j n)$, the bias term is small when λ_1/λ_j is not large.

5.2. Test Procedure Using Eigenstructures

Let $\tilde{x}_{ijl} = \tilde{\boldsymbol{h}}_{ijl}^T\boldsymbol{x}_{il}$ for all i, j, l, where the $\tilde{\boldsymbol{h}}_{ijl}$'s are defined by (5.4). From Propositions 3 and 5, we consider the test statistic for (1.1),

$$\begin{aligned} \widehat{T}_* =& 2\sum_{i=1}^{2} \frac{\sum_{l<l'}^{n_i}(\boldsymbol{x}_{il}^T\boldsymbol{x}_{il'} - \sum_{j=1}^{k_i}\tilde{x}_{ijl}\tilde{x}_{ijl'})}{n_i(n_i-1)} \\ &- 2\frac{\sum_{l=1}^{n_1}\sum_{l'=1}^{n_2}(\boldsymbol{x}_{1l} - \sum_{j=1}^{k_1}\tilde{x}_{1jl}\tilde{\boldsymbol{h}}_{1j})^T(\boldsymbol{x}_{2l'} - \sum_{j=1}^{k_2}\tilde{x}_{2jl'}\tilde{\boldsymbol{h}}_{2j})}{n_1 n_2}, \end{aligned}$$

where the $\tilde{\boldsymbol{h}}_{ij}$'s are defined by (5.3). We assume the following conditions when (A-vi) is met.

(A-ix) $\dfrac{\lambda_{i1}^2}{n_i\Psi_{i(k_i+1)}} \to 0$ as $m \to \infty$ for $i = 1, 2$;

(A-x) $\dfrac{\boldsymbol{\mu}_{1*}^T\boldsymbol{\Sigma}_{i*}\boldsymbol{\mu}_{1*}+\boldsymbol{\mu}_{2*}^T\boldsymbol{\Sigma}_{i*}\boldsymbol{\mu}_{2*}}{\Psi_{i(k_i+1)}} \to 0$ as $p\to\infty$ and

$$\limsup_{m\to\infty}\frac{n_i\{\mu_{i(j)}^2+(\boldsymbol{h}_{ij}^T\boldsymbol{\mu}_{i'*})^2\}}{\lambda_{ij}}<\infty\ (i'\neq i)\ \text{ for } i=1,2;\ j=1,...,k_i.$$

Then, we have the following result.

Theorem 5. *If (A-vi) and (A-viii) to (A-x) hold, then* $\widehat{T}_*-T_*=o_P(K_{1*}^{1/2})$ *as* $m\to\infty$. *If also* $\limsup_{m\to\infty}\Delta_*^2/K_{1*}<\infty$, *then* $(\widehat{T}_*-\Delta_*)/K_*^{1/2}\Rightarrow N(0,1)$ *as* $m\to\infty$.

By using Lemma 1, $K_{1*}/K_*=1+o(1)$ as $m\to\infty$ under (A-vi) and $\limsup_{m\to\infty}\Delta_*^2/K_{1*}<\infty$. Thus, we consider estimating K_{1*}. Let $\widehat{\boldsymbol{A}}_{i(k_i)}=\boldsymbol{I}_p-\sum_{j=1}^{k_i}\hat{\boldsymbol{h}}_{ij}\hat{\boldsymbol{h}}_{ij}^T$ for $i=1,2$. We estimate K_{1*} by

$$\widehat{K}_{1*}=2\sum_{i=1}^{2}\frac{\widehat{\Psi}_{i(k_i+1)}}{n_i(n_i-1)}+4\frac{\mathrm{tr}(\boldsymbol{S}_{1n_1}\widehat{\boldsymbol{A}}_{1(k_1)}\boldsymbol{S}_{2n_2}\widehat{\boldsymbol{A}}_{2(k_2)})}{n_1n_2},$$

where $\widehat{\Psi}_{i(k_i+1)}$ is defined by (S2.1) of the supplementary material.

Lemma 3. *If (A-vi), (A-viii) and (A-ix) hold, then* $\widehat{K}_{1*}/K_{1*}=1+o_P(1)$ *as* $m\to\infty$.

Now, we test (1.1) by

$$\text{rejecting } H_0 \Longleftrightarrow \widehat{T}_*/\widehat{K}_{1*}^{1/2}>z_\alpha. \tag{5.5}$$

Let power(Δ_*) denote the power of the test (5.5). Then, from Theorem 5 and Lemma 3, we have the following result.

Theorem 6. *If (A-vi) and (A-vii) to (A-x) hold, then the test (5.5) has,*

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

as $m \to \infty$,

$$\textit{size} = \alpha + o(1) \quad \textit{and} \quad \textit{power}(\Delta_*) - \Phi\bigg(\frac{\Delta_*}{K_*^{1/2}} - z_\alpha\Big(\frac{K_{1*}}{K_*}\Big)^{1/2}\bigg) = o(1).$$

In general, the k_i's are unknown in $\widehat{T}_*$ and $\widehat{K}_{1*}$. See Section S2.2 in the supplementary material for estimation of the k_i's. If (4.1) is met, one may use the test (4.2). However, under (4.1), (A-vi) and $\limsup_{m\to\infty} \Delta_*^2/K_{1*} < \infty$, we note that $\mathrm{Var}(T_*)/\mathrm{Var}(T_I) = O(K_{1*}/K_1) \to 0$ as $m \to \infty$, so that the power of (4.2) must be lower than that of (5.5). See Section 6 for numerical comparisons. We recommend the use of the test (5.5) for the SSE model in general.

5.3. How to Check SSE Models and Estimate Parameters

We provide a method to distinguish between the NSSE model at (1.4) and the SSE model at (1.6). We also give a method to estimate the parameters required in the test procedure (5.5). We summarize the results in Section S2 of the supplementary material.

5.4. Demonstration

We introduce two high-dimensional data sets that obey the SSE model. We illustrate the proposed test procedure at (5.5) by using microarray data sets. We summarize the results in Section S3 of the supplementary material.

6. Simulations for Strongly Spiked Eigenvalue Model

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

We used computer simulations to study the performance of the test procedures at (4.2) and (5.5) for the SSE model. In general, the k_i's are unknown for (5.5). Hence, we estimated k_i by $\hat{k}_i$, where $\hat{k}_i$ is given in Section S2.2 of the supplementary material. We set $\kappa(n_i) = (n_i^{-1}\log n_i)^{1/2}$ in (S2.2) of the supplementary material. We checked the performance of the test procedure at (5.5) with $k_i = \hat{k}_i$, $i = 1, 2$. We considered a naive estimator of T_* as $T(\widehat{\boldsymbol{A}}_{1(k_1)}, \widehat{\boldsymbol{A}}_{2(k_2)})$ and checked the performance of the test procedure given by

$$\text{rejecting } H_0 \Longleftrightarrow T(\widehat{\boldsymbol{A}}_{1(k_1)}, \widehat{\boldsymbol{A}}_{2(k_2)})/\widehat{K}_{1*}^{1/2} > z_\alpha. \tag{6.1}$$

We also checked the performance of the test procedure at (3.1) with $\boldsymbol{A} = \boldsymbol{I}_p$. We set $\alpha = 0.05$, $\boldsymbol{\mu}_1 = \boldsymbol{0}$, and

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{(1)} & \boldsymbol{O}_{2,p-2} \\ \boldsymbol{O}_{p-2,2} & c_i\boldsymbol{\Sigma}_{(2)} \end{pmatrix} \text{ with } \boldsymbol{\Sigma}_{(1)} = \mathrm{diag}(p^{2/3}, p^{1/2}) \text{ and } \boldsymbol{\Sigma}_{(2)} = (0.3^{|i-j|^{1/2}})$$

for $i = 1, 2$, where $\boldsymbol{O}_{l,l'}$ is the $l \times l'$ zero matrix and $(c_1, c_2) = (1, 1.5)$. Here (4.1) and (A-vi) with $k_1 = k_2 = 2$ are met. When considering the alternative hypothesis, we set $\boldsymbol{\mu}_2 = (0, ..., 0, 1, 1, 1, 1)^T$ with last four elements 1. We considered three cases:

(a) $\pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $p = 2^s$, $n_1 = 3\lceil p^{1/2}\rceil$ and $n_2 = 4\lceil p^{1/2}\rceil$ for $s = 4, ..., 10$;

(b) The $\boldsymbol{z}_{ij}$'s are i.i.d. as the p-variate t-distribution, $t_p(\nu)$, with mean zero, covariance matrix $\boldsymbol{I}_p$, and degrees of freedom $\nu = 15$, $(n_1, n_2) = (40, 60)$, and $p = 50 + 100(s-1)$ for $s = 1, ..., 7$;

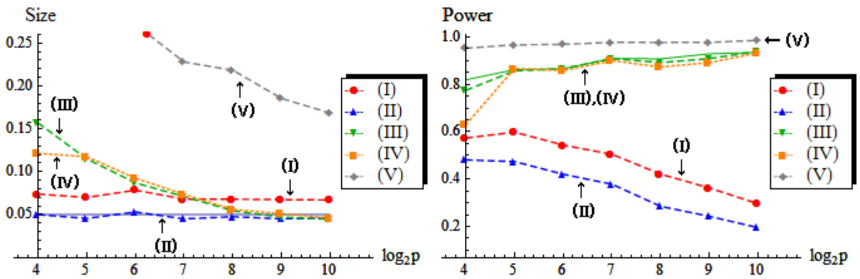
(c) $z_{itj} = (v_{itj} - 5)/10^{1/2}$ $(t = 1, ..., p)$ in which the v_{itj}'s are i.i.d. as χ^2_5, $p = 500$, $n_1 = 10s$ and $n_2 = 1.5n_1$ for $s = 2, ..., 8$.

Here (A-viii) is met both for (a) and (c). However, (A-viii) (or (A-i)) is not met for (b). Similar to Section 3.3, we calculated $\overline{\alpha}$ and $1 - \overline{\beta}$ with 2000 replications for five test procedures: (I) from (3.1) with $\boldsymbol{A} = \boldsymbol{I}_p$, (II) from (4.2), (III) from (5.5), (IV) from (5.5) with $k_i = \hat{k}_i$, $i = 1, 2$, and (V) from (6.1). Their standard deviations are less than 0.011. In Fig. 2, for (a) to (c), we plotted $\overline{\alpha}$ in the left panel and $1 - \overline{\beta}$ in the right panel. From Theorem 6, we plotted the asymptotic power, $\Phi(\Delta_*/K_*^{1/2} - z_\alpha(K_{1*}/K_*)^{1/2})$, for (III).

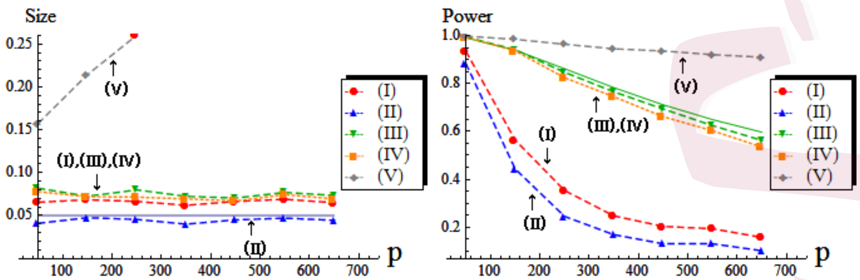
We observe that (II) gives better performances compared to (I) regarding size. The size of (I) did not get close to α, probably because T_I does not satisfy the asymptotic normality given in Theorem 2 when (1.4) is not met. On the other hand, (II) (or (I)) gave quite poor performances compared to (III) and (IV) regarding power, probably because $\text{Var}(T_I)/\text{Var}(T_*) \to \infty$ as $p \to \infty$ in the current setting. The size of (V) was much higher than α, probably because of the bias of $T(\widehat{\boldsymbol{A}}_{1(k_1)}, \widehat{\boldsymbol{A}}_{2(k_2)})$. See Section 5.1 for the details. We observe that (III) and (IV) gave adequate performances even in the non-Gaussian cases. The performances of (III) and (IV) were similar to each other in almost all cases. When p and the n_i's are not small, the plots of (IV) were close to the theoretical values. Hence, we recommend the use of the test procedure at (5.5) with $k_i = \hat{k}_i$, $i = 1, 2$, when (1.6) holds.

We also checked the performance of the test procedures for the MSN distribution and the multivariate skew t (MST) distribution. See Azzalini

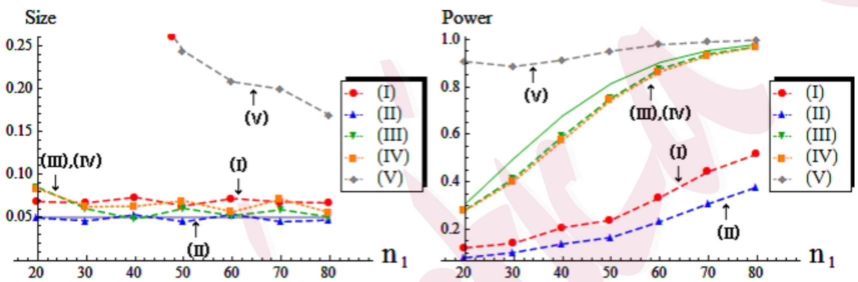
Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063



(a) $\pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $p = 2^s$, $n_1 = 3\lceil p^{1/2}\rceil$ and $n_2 = 4\lceil p^{1/2}\rceil$ for $s = 4, ..., 10$.



(b) $\boldsymbol{z}_{ij}$s are i.i.d. as $t_p(15)$, $(n_1, n_2) = (40, 60)$ and $p = 50 + 100(s - 1)$ for $s = 1, ..., 7$.



(c) $z_{irj} = (v_{itj} - 5)/10^{1/2}$ $(t = 1, ..., p)$ in which v_{itj}s are i.i.d. as χ^2_5, $p = 500$, $n_1 = 10s$ and $n_2 = 1.5n_1$ for $s = 2, ..., 8$.

Figure 2: The performances of five tests: (I) from (3.1) with $\boldsymbol{A} = \boldsymbol{I}_p$, (II) from (4.2), (III) from (5.5), (IV) from (5.5) with $k_i = \hat{k}_i$, $i = 1, 2$, and (V) from (6.1). For (a) to (c), the values of $\overline{\alpha}$ are denoted by the dashed lines in the left panel and the values of $1 - \overline{\beta}$ are denoted by the dashed lines in the right panel. The asymptotic power of (III) was given by $\Phi(\Delta_*/K_*^{1/2} - z_\alpha(K_{1*}/K_*)^{1/2})$ which is denoted by the solid line in the right panels. When n_is are small or p is large, $\overline{\alpha}$ for (V) was too high to describe.

and Capitanio (2003) and Gupta (2003) for the details of the MST distribution. We give the results in Section S4.2 of the supplementary material.

7. Conclusion

By classifying eigenstructures into two classes, the SSE and NSSE models, and then selecting a suitable test procedure depending on the eigenstructure, we can quickly obtain a much more accurate result at lower computational cost. These benefits are vital in groundbreaking research of medical diagnostics, engineering, big data analysis, etc.

Supplementary Materials

We give data analyses and proofs of the theoretical results, together with additional simulations, in the online supplementary material. We also give methods to distinguish between the NSSE model and the SSE model, and estimate the parameters required in the test procedure (5.5).

Acknowledgements

We thank an associate editor and two anonymous referees for their constructive comments. The authors are grateful to the Co-Editor, Hsin-Cheng Huang, for his helpful remarks. Research of the first author was partially supported by Grants-in-Aid for Scientific Research (A) and Challenging Exploratory Research, Japan Society for the Promotion of Science (JSPS), under Contract Numbers 15H01678 and 26540010. Research of the second author was partially supported by Grant-in-Aid for Young Scientists (B), JSPS, under Contract Number 26800078.

References

Aoshima, M. and Yata, K. (2011). Two-stage procedures for high-dimensional data. *Sequential Anal. (Editor's special invited paper)* **30**, 356-399.

Aoshima, M. and Yata, K. (2015). Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. *Methodol. Comput. Appl. Probab.* **17**, 419-439.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Statist. Soc. B* **65**, 367-389.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6**, 311-329.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36**, 2577-2604.

Cai, T. T., Liu, W., and Xia, Y. (2014). Two sample test of high dimensional means under dependence. *J. R. Statist. Soc. B* **76**, 349-372.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38**, 808-835.

Chen, S. X., Zhang, L.-X., and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105**, 810-819.

Dempster, A. P. (1958). A high dimensional two sample significance test. *Ann. Math. Statist.* **29**, 995-1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics* **16**, 41-50.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Statist. Soc. B* **75**, 603-680.

Gupta, A. K. (2003). Multivariate skew t-distribution. *Statistics* **37**, 359-363.

Ishii, A., Yata, K., and Aoshima, M. (2016). Asymptotic properties of the first principal component and equality tests of covariance matrices in high-dimension, low-sample-size context. *J. Statist. Plan. Infer.* **170**, 186-199.

Jung, S. and Marron, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37**, 4104-4130.

Katayama, S., Kano, Y., and Srivastava, M. S. (2013). Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **116**, 410-421.

Ma, Y., Lan, W., and Wang, H. (2015). A high dimensional two-sample test under a low dimensional factor structure. *J. Multivariate Anal.* **140**, 162-170.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Econometrics* **168**, 244-258.

Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *J. Japan Statist. Soc.* **37**, 53-86.

Srivastava, M. S., Katayama, S., and Kano, Y. (2013). A two sample test in high dimensional data. *J. Multivariate Anal.* **114**, 349-358.

Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.* **105**, 193-215.

Yata, K. and Aoshima, M. (2013a). Correlation tests for high-dimensional data using extended cross-data-matrix methodology. *J. Multivariate Anal.* **117**, 313-331.

Yata, K. and Aoshima, M. (2013b). PCA consistency for the power spiked model in high-dimensional settings. *J. Multivariate Anal.* **122**, 334-354.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0063

Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan.

E-mail: aoshima@math.tsukuba.ac.jp

Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan.

E-mail: yata@math.tsukuba.ac.jp