Statistica Sinica Preprint No: SS-2022-0377	
Title	Invariance Principle and Clt for the Spiked Eigenvalues of Large-dimensional Fisher Matrices and Applications
Manuscript ID	SS-2022-0377
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0377
Complete List of Authors	Dandan Jiang, Zhiqiang Hou, Zhidong Bai and Runze Li
Corresponding Authors	Runze Li
E-mails	rzli@psu.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

Statistica Sinica

Invariance principle and CLT for the spiked eigenvalues of large-dimensional Fisher matrices and applications

Dandan Jiang[1], Zhiqiang Hou[2], Zhidong Bai[3] and Runze Li[4]

[1]*Xi'an Jiaotong University,* [2]*Shandong University of Finance and Economics,*

[3]*Northeast Normal University and* [4]*Pennsylvania State University at University Park*

Abstract: This paper aims to derive the asymptotic distributions of the spiked eigenvalues of large-dimensional spiked Fisher matrices, without imposing Gaussian assumptions or restrictive assumptions on covariance matrices. We first establish an invariance principle for the spiked eigenvalues of the Fisher matrix. That is, we show that the limiting distributions of the spiked eigenvalues are invariant over a broad range of population distributions satisfying certain conditions. Utilizing this invariance principle, we establish a central limit theorem (CLT) for the spiked eigenvalues, and further explore some interesting applications by using the CLT to derive the power functions of the Roy Maximum root test for linear hypotheses in linear models, as well as the test in signal detection. To evaluate the effectiveness of the newly proposed test, we conduct Monte Carlo simulation studies and compare its performance with existing tests.

Key words and phrases: Roy maximum root test, Random matrix theory, Spiked model, Two-sample covariance problem.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

1. Introduction

Motivated by several applications of hypothesis on two-sample covariance matrices and linear hypothesis on regression coefficient matrix in linear models, we consider the following two-sample spiked model. Let $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ be the covariance matrices from two p-dimensional populations, and let $\mathbf{S}_1$ and $\mathbf{S}_2$ be the corresponding sample covariance matrices with sample sizes n_1 and n_2. The two-sample spiked model assumes that $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1 + \mathbf{\Delta}$, where $\mathbf{\Delta}$ is a $p \times p$ matrix of finite rank M. It is of great interest to study statistical inference on the spikes, including, but not limited to, testing the presence of the spikes, testing the number of the spikes, and calculating the power under the alternative hypothesis in two-sample testing problems. Thus, it is critical to establish the asymptotic properties of the spiked eigenvalues of a Fisher matrix $\mathbf{F} = \mathbf{S}_1 \mathbf{S}_2^{-1}$. It is of particular interest to derive the asymptotic distribution of $\lambda_{\max}(\mathbf{F})$, the largest eigenvalue of $\mathbf{F}$. However, the existing related works are limited due to imposed strict conditions such as the blockwise diagonal assumption on $\mathbf{\Sigma}_1 \mathbf{\Sigma}_2^{-1}$, rank-one perturbation and etc. These limitations narrow the scope of applications, which will be elaborated later on. Therefore it is necessary to investigate the limiting behavior of the eigenvalues of the large-dimensional Fisher matrix under less strict assumptions.

To extend the existing work to a broader spectrum, we establish an invariance principle for the spiked eigenvalues of the Fisher matrix, and the invariance principle can be used as a universal probability tool for deriving the asymptotic distribution of local spectral statistics of $\mathbf{F}$ under mild assumptions on $\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}$. Before we proceed further, let us briefly review some inspiring studies on the one-sample spiked model. The one-sample spiked model $\mathbf{\Sigma} = \mathbf{I}_p + \mathbf{\Delta}$ has received a lot of attentions in the literatures, where $\mathbf{I}_p$ is the identity matrix, and $\mathbf{\Delta}$ is a low-rank $p \times p$ matrix. Johnstone (2001) was the first work proposing one-sample spiked model to study principal component analysis (PCA). Since the work by Johnstone (2001), many works have been published and focus on studying the asymptotic law for spiked eigenvalues of large-dimensional covariance matrix (Baik et al., 2005; Baik and Silverstein, 2006; Paul, 2007; Bai and Yao, 2008). Also see (Bai and Yao, 2012; Fan and Wang, 2017; Cai et al., 2020; Jiang and Bai, 2021) for a more general one-sample spiked model. These works establish the limiting distribution for the spiked eigenvalues of the sample covariance matrices under different settings.

Compared with the one-sample spiked model, there are relatively few studies on two-sample spiked model. Zheng et al. (2017) derived the limiting distribution of the eigenvalues of a Fisher matrix and established a CLT for a wide class of functions of all the eigenvalues as a whole. According to Han et al. (2016), the largest eigenvalue of the Fisher matrix follows the Tracy-Widom law under some conditions. Therefore, the results in Zheng et al. (2017) are not applicable for the local spectral statistics, especially for those made of spiked eigenvalues. For a simplified two-sample spiked models assuming that $\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}$ is a rank M perturbation of identity matrix with diagonal independence and bounded population fourth moment, Wang and Yao (2017) established CLT for the extreme eigenvalues of large-dimensional spiked Fisher matrices. Based on Wang and Yao (2017), Xie et al (2021) investigated the limiting laws for extreme eigenvalues of the Fisher matrix when the number of spikes is divergent and these spikes are unbounded under the diagonal $\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}$ assumption. Johnstone and Onatski (2020) proposed the rank-one two-sample spiked models that represent the James' classes in James (1964), and further derived the asymptotic behavior of the likelihood ratios in large-dimensional setting. The aforementioned works impose some restrictive or unrealistic conditions such as only one threshold, diagonal assumption, rank-one perturbation, etc. The first two conditions (i.e., only one threshold and diagonal assumption) imply that the spiked eigenvalues and non-spiked eigenvalues are generated by independent variables. The assumption of rank-one perturbation means that

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

there is only one spike, which is just an extremely special case.

In this paper, we study the asymptotic properties of Fisher matrix of a two-sample spiked model under a new setting, which allows that (a) two populations may have different general covariance matrices, (b) the spiked eigenvalues of the Fisher matrix may be scattered into the spaces of a few groups, and (c) the largest eigenvalue may tend to infinity. Under this new setting, the fourth moments of population are not required to be bounded. For ease of presentation, we refer to a Fisher matrix with this new setting as a *generalized spiked Fisher matrix*. Under a general setting, we establish an invariance principle for the generalized spiked Fisher matrix by using a similar but more complicated technique of Jiang and Bai (2021). It is worth noting that Jiang and Bai (2021) focused on the one-sample spiked model with only one random covariance matrix, which cannot cover the spectral statistics of a Fisher matrix with the product of two random covariance matrices without an explicit expression. Compared with the existing works on spiked Fisher matrices, this work relaxes the bounded fourth-moment condition on population to a tail probability condition, which is a regular and necessary condition for the weak convergence of the largest eigenvalue. With the aid of the invariance principle theorem, we further establish the CLT for the local spiked eigenvalues of large-dimensional generalized spiked

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

Fisher matrices under mild assumptions on the population distribution. As applications, we use the CLT to derive the power functions of Roy maximum root test on linear hypothesis in large-dimensional linear models and the signal detection test. As a by-product, our results naturally extend the result of Wang and Yao (2017) to a general case under which we can successfully remove the diagonal or block-wise diagonal assumption on the matrix $\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}$. Our setting allows the spiked eigenvalues to be generated from variables that are partially dependent on the ones corresponding to the non-spiked eigenvalues. Additionally, our setting allows for the use of multiple pairs of thresholds for groups of spiked eigenvalues. In summary, our setting is more realistic than those in the existing works.

The rest of this paper is organized as follows. We establish the invariance principle and the CLT of generalized spiked Fisher matrix in Sections 2 and 3, respectively. We use the CLT to derive the local power functions of Roy maximum root test for linear hypothesis in large-dimensional linear models and the test in signal detection in Section 4. We present some numerical study in Section 5. Additional technical proofs are presented in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

2. Invariance principle for generalized spiked Fisher matrix

2.1 Phase transition for the spiked eigenvalues

Assume that

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_1}) = (x_{ij}), 1 \leq i \leq p,\ 1 \leq j \leq p,$$

$$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n_2}) = (y_{kl}), 1 \leq k \leq p,\ 1 \leq l \leq p,$$

are two p-dimensional arrays with components having zero mean and identity variance. Denote $\mathbf{\Sigma}_1^{1/2}\mathbf{X}$ and $\mathbf{\Sigma}_2^{1/2}\mathbf{Y}$ to be two independent sample matrices, where $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are two nonnegative definite matrices. We denote the corresponding sample covariance matrices of the two observations:

$$\mathbf{S}_1 = \frac{1}{n_1}\mathbf{\Sigma}_1^{1/2}\mathbf{X}\mathbf{X}^*\mathbf{\Sigma}_1^{1/2}, \quad \mathbf{S}_2 = \frac{1}{n_2}\mathbf{\Sigma}_2^{1/2}\mathbf{Y}\mathbf{Y}^*\mathbf{\Sigma}_2^{1/2}. \tag{2.1}$$

The matrix $\mathbf{F} = \mathbf{S}_1\mathbf{S}_2^{-1}$ is so-called generalized Fisher matrix in Jiang et al. (2021), where the condition $p < n_2$ ensures that $\mathbf{S}_2$ is invertible. The eigenvalues of $\mathbf{F}$ are denoted by

$$l_{p,1} \geq \cdots \geq l_{p,j} \geq \cdots \geq l_{p,p}. \tag{2.2}$$

It is well known that the eigenvalues of $\mathbf{F}$ are the same as those of the matrix with the form below (Still use $\mathbf{F}$ for brevity, if no confusion):

$$\mathbf{F} = \mathbf{T}_p^*\tilde{\mathbf{S}}_1\mathbf{T}_p\tilde{\mathbf{S}}_2^{-1}, \tag{2.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

where $\mathbf{T}_p = \mathbf{\Sigma}_1^{1/2}\mathbf{\Sigma}_2^{-1/2}$, $\tilde{\mathbf{S}}_1 = (1/n_1)\mathbf{X}\mathbf{X}^*$ and $\tilde{\mathbf{S}}_2 = (1/n_2)\mathbf{Y}\mathbf{Y}^*$. Assume that the spectrum of $\mathbf{T}_p^*\mathbf{T}_p$ is formed as

$$\beta_{p,1}, \cdots, \beta_{p,j}, \cdots, \beta_{p,p} \tag{2.4}$$

in descending order. Moreover, the singular value decomposition of $\mathbf{T}_p$ is defined as

$$\mathbf{T}_p = \mathbf{U} \begin{pmatrix} \mathbf{D}_1^{\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2^{\frac{1}{2}} \end{pmatrix} \mathbf{V}^*, \tag{2.5}$$

where $\mathbf{D}_1$ is a diagonal matrix of the M spiked eigenvalues of the generalized spiked Fisher matrix $\mathbf{F}$, and $\mathbf{D}_2$ is the diagonal matrix of the non-spiked ones with bounded components. Moreover, $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices (orthogonal matrices for real case). Let $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$, and $\mathbf{U}_1$ and $\mathbf{U}_2$ are two matrices with size $p \times M$ and $p \times (p - M)$. Same consideration applies to $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$.

To derive the limiting law for the spiked eigenvalues of $\mathbf{F}$, we need to introduce some notation and preliminary assumptions. For any $n \times n$ matrix $\mathbf{A}_n$ with real eigenvalues, let F_n be the *empirical spectral distribution* (ESD) function of $\mathbf{A}_n$, which is defined as:

$$F_n(x) = \frac{1}{n} Card\{i; \lambda_i^{\mathbf{A}_n} \leq x\},$$

where $\lambda_i^{\mathbf{A}_n}$ denotes the i-th largest eigenvalue of $\mathbf{A}_n$. If F_n has a limiting distribution F, then we call it the *limiting spectral distribution* (LSD) of

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

sequence $\{\mathbf{A}_n\}$. For any function of bounded variation G on the real line, its *Stieltjes transform* is defined by

$$m(z) = \int \frac{1}{\lambda - z} dG(\lambda), \quad z \in \mathbb{C}^+.$$

Assumption 1. The two double arrays $\{x_{ij}, i, j = 1, 2, ...\}$ and $\{y_{ij}, i, j = 1, 2, ...\}$ consist of independent and identically distributed (i.i.d.) random variables with a mean of zero and a variance of one. Furthermore, it holds that $\mathrm{E}x_{ij}^2 = 0$ and $\mathrm{E}y_{ij}^2 = 0$ in the complex case if the variables and $\mathbf{T}_p$ are complex.

Assumption 2. Assuming that $c_{n_1} = p/n_1 \in (0, \infty)$, $c_{n_2} = p/n_2 \in (0, 1)$ is considered throughout the paper when $\min(p, n_1, n_2) \to \infty$.

Assumption 3. Assuming that $\mathbf{T}_p = \mathbf{\Sigma}_1^{1/2}\mathbf{\Sigma}_2^{-1/2}$ is nonrandom, and the ESD $H_n(t)$ of $\{\mathbf{T}_p^*\mathbf{T}_p\}$ satisfies $H_n \xrightarrow{w} H$, ($\xrightarrow{w}$ denotes weak convergence), where H is a nonrandom probability measure.

Assumption 4. Assume that

$$\lim_{\tau\to\infty} \tau^4 \mathrm{P}\left(|x_{11}| > \tau\right) = 0 \text{ and } \lim_{\tau\to\infty} \tau^4 \mathrm{P}\left(|y_{11}| > \tau\right) = 0 \tag{2.6}$$

hold for the i.i.d. samples, where both of the fourth moments are not necessarily required to exist.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

Assumption 5. Suppose that

$$\max_{t,s}|u_{ts}|^2\left[\mathrm{E}\{|x_{11}|^4\delta(|x_{11}|<\eta_{n_1}\sqrt{n_1})\}-2-q\right]\to 0,$$
$$\max_{t,s}|v_{ts}|^2\left[\mathrm{E}\{|y_{11}|^4\delta(|y_{11}|<\eta_{n_2}\sqrt{n_2})\}-2-q\right]\to 0,$$

where $q=1$ for real case and $q=0$ for complex, $\delta(\cdot)$ is the indicator function, $(u_{ts})=\mathbf{U}_1$, and $(v_{ts})=\mathbf{V}_1$. Here η_{n_1} and η_{n_2} are two constants sequences concerning n_1 and n_2 converging to 0.

These assumptions are interpreted as follows: Assumptions 1–2 are the commonly-used condition imposed in large-dimensional settings, whereas the dependent relationship are reflected by the matrix $\mathbf{T}_p$. Following the Assumption 3, the spiked eigenvalues of $\{\mathbf{T}_p^*\mathbf{T}_p\}$ are subject to the condition

$$\beta_{p,j_k+1}=\beta_{p,j_k+2}=\cdots=\beta_{p,j_k+m_k}\overset{def}{=}\alpha_k,\quad k\in\{1,\ldots,K\}. \tag{2.7}$$

Here α_k's are outside the support of H and satisfy the separation condition, which means

$$\min_{k\neq j}\left|\frac{\alpha_k}{\alpha_j}-1\right|>d, \tag{2.8}$$

where d is a positive constant that is independent of n. In addition, $\mathcal{J}_k=\{j_k+1,\ldots,j_k+m_k\}$ denotes the set of ranks of α_k, where m_k represents the multiplicities of α_k that satisfy $m_1+\cdots+m_K=M$, a fixed integer. We call $\alpha_k, k\in\{1,...,K\}$ defined in (2.7) spiked eigenvalues. Note that α_k can be

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

located in any gap between disconnected sections of the support of H, which means that α_k is not just the extreme eigenvalues of $\{\mathbf{T}_p^*\mathbf{T}_p\}$. Under all of Assumptions 1–3, the limits of the eigenvalues $l_{p,i}, i \in \{1, \ldots, p\}$ associated to spiked eigenvalues α_k, $k \in \{1, \ldots, K\}$, are derived in Jiang et al. (2021) when x_{ij} and y_{ij} have bounded fourth moments. Details are presented in the following proposition due to Jiang et al. (2021). Define the function ψ_k as follows.

$$\psi_k := \psi(\alpha_k) = \frac{\alpha_k \left\{1 - c_1 \int t/(t-\alpha_k)\mathrm{d}H(t)\right\}}{1 + c_2 \int \alpha_k/(t-\alpha_k)\mathrm{d}H(t)}. \tag{2.9}$$

Proposition 1. *For any spiked eigenvalue α_k with multiplicity m_k, for $k = 1, \cdots, K$, of $\mathbf{F}$ defined in (2.3), let*

$$\rho_k = \begin{cases} \psi(\alpha_k), & \text{if } \psi'(\alpha_k) > 0; \\ \psi(\underline{\alpha}_k), & \text{if there exists } \underline{\alpha}_k \text{ such that } \psi'(\underline{\alpha}_k) = 0, \\ & \text{and } \psi'(t) < 0, \text{ for all } \alpha_k \le t < \underline{\alpha}_k; \\ \psi(\overline{\alpha}_k), & \text{if there exists } \overline{\alpha}_k \text{ such that } \psi'(\overline{\alpha}_k) = 0, \\ & \text{and } \psi'(s) < 0, \text{ for all } \overline{\alpha}_k < s \le \alpha_k. \end{cases}$$

Then, under Assumptions 1–3 with the bounded fourth moments of x_{ij} and y_{ij}, the sequence $\{l_{p,j}/\rho_k - 1,\ j \in \mathcal{J}_k\}$ converges almost surely to 0.

Since the convergence of $c_{n_1} \to c_1$, $c_{n_2} \to c_2$ and $H_n(t) \to H(t)$ can be very slow, the difference $\sqrt{n}(l_{p,j} - \psi_k)$ may not have a limiting distribution.

Thus, we use

$$\psi_{n,k} := \psi_n(\alpha_k) = \frac{\alpha_k \left\{1 - c_{n_1} \int t/(t-\alpha_k) \mathrm{d}H_n(t)\right\}}{1 + c_{n_2} \int \alpha_k/(t-\alpha_k) \mathrm{d}H_n(t)}$$

instead of ψ_k in ρ_k, and n denotes (n_1, n_2), particularly for deriving the CLT in the following section.

Note that the Proposition 1 holds under the bounded fourth-moment assumption. To relax this assumption, Assumption 4 on the tail probability is provided in (2.6). We then conduct truncation and centralization procedures for the variables as in Silverstein (1995) and prove that the Proposition 1 still holds in probability without the bounded fourth-moment assumption if Assumption 4 is satisfied. The details are deferred to the Supplement Material. Assumption 5 is a technical condition used for the proof of universality. The detailed explanation of Assumption 5 can be found in the Supplement A in Jiang and Bai (2021).

Based on these assumptions, an invariance principle theorem is established in Section 2.2. The theorem guarantees the universality of the limiting distribution of spiked eigenvalues from a generalized spiked Fisher matrix.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

2.2 Invariance principle theorem

For the generalized spiked Fisher matrix $\mathbf{F} = \mathbf{T}_p^*\tilde{\mathbf{S}}_1\mathbf{T}_p\tilde{\mathbf{S}}_2^{-1}$ defined in (2.3), we consider the arbitrary sample spiked eigenvalue of $\mathbf{F}$, $l_{p,j}, j \in \mathcal{J}_k, k = 1, \ldots, K$. By the singular value decomposition of $\mathbf{T}_p$ in (2.5) and the eigen-equation for $\mathbf{F}$, we have

$$0 = |l_{p,j}\mathbf{I} - \mathbf{F}| = \left| l_{p,j}\mathbf{I} - \mathbf{V}\begin{pmatrix} \mathbf{D}_1^{\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2^{\frac{1}{2}} \end{pmatrix} \mathbf{U}^*\tilde{\mathbf{S}}_1\mathbf{U} \begin{pmatrix} \mathbf{D}_1^{\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2^{\frac{1}{2}} \end{pmatrix} \mathbf{V}^*\tilde{\mathbf{S}}_2^{-1} \right|.$$

It is equivalent to

$$\begin{aligned} 0 &= |l_{p,j}\mathbf{V}^*\tilde{\mathbf{S}}_2\mathbf{V} - \mathrm{diag}(\mathbf{D}_1^{\frac{1}{2}}, \mathbf{D}_2^{\frac{1}{2}})\mathbf{U}^*\tilde{\mathbf{S}}_1\mathbf{U}\mathrm{diag}(\mathbf{D}_1^{\frac{1}{2}}, \mathbf{D}_2^{\frac{1}{2}})| \\ &= \left| \begin{pmatrix} l_{p,j}\mathbf{V}_1^*\tilde{\mathbf{S}}_2\mathbf{V}_1, & l_{p,j}\mathbf{V}_1^*\tilde{\mathbf{S}}_2\mathbf{V}_2 \\ l_{p,j}\mathbf{V}_2^*\tilde{\mathbf{S}}_2\mathbf{V}_1, & l_{p,j}\mathbf{V}_2^*\tilde{\mathbf{S}}_2\mathbf{V}_2 \end{pmatrix} - \begin{pmatrix} \mathbf{D}_1^{\frac{1}{2}}\mathbf{U}_1^*\tilde{\mathbf{S}}_1\mathbf{U}_1\mathbf{D}_1^{\frac{1}{2}}, & \mathbf{D}_1^{\frac{1}{2}}\mathbf{U}_1^*\tilde{\mathbf{S}}_1\mathbf{U}_2\mathbf{D}_2^{\frac{1}{2}} \\ \mathbf{D}_2^{\frac{1}{2}}\mathbf{U}_2^*\tilde{\mathbf{S}}_1\mathbf{U}_1\mathbf{D}_1^{\frac{1}{2}}, & \mathbf{D}_2^{\frac{1}{2}}\mathbf{U}_2^*\tilde{\mathbf{S}}_1\mathbf{U}_2\mathbf{D}_2^{\frac{1}{2}} \end{pmatrix} \right|. \end{aligned}$$

By Schur formula, if $l_{p,j}$ is a sample spiked eigenvalue of $\mathbf{F}$, then it is not of $\mathbf{D}_2^{\frac{1}{2}}\mathbf{U}_2^*\tilde{\mathbf{S}}_1\mathbf{U}_2\mathbf{D}_2^{\frac{1}{2}}\left(\mathbf{V}_2^*\tilde{\mathbf{S}}_2\mathbf{V}_2\right)^{-1}$. To derive the fluctuation of $l_{p,j}, j \in \mathcal{J}_k$, we will technically disassemble the determinant into four parts, making each part manageable. The following equation holds for every sampled spiked eigenvalue, $l_{p,j}, j \in \mathcal{J}_k$, for $k = 1, \cdots, K$,

$$\begin{aligned} 0 = & \Big| l_{p,j}\mathbf{V}_1^*\tilde{\mathbf{S}}_2\mathbf{V}_1 - \mathbf{D}_1^{\frac{1}{2}}\mathbf{U}_1^*\tilde{\mathbf{S}}_1\mathbf{U}_1\mathbf{D}_1^{\frac{1}{2}} - (l_{p,j}\mathbf{V}_1^*\tilde{\mathbf{S}}_2\mathbf{V}_2 - \mathbf{D}_1^{\frac{1}{2}}\mathbf{U}_1^*\tilde{\mathbf{S}}_1\mathbf{U}_2\mathbf{D}_2^{\frac{1}{2}})\mathbf{Q}^{-\frac{1}{2}} \\ & \cdot (l_{p,j}\mathbf{I} - \mathbf{Q}^{-\frac{1}{2}}\mathbf{D}_2^{\frac{1}{2}}\mathbf{U}_2^*\tilde{\mathbf{S}}_1\mathbf{U}_2\mathbf{D}_2^{\frac{1}{2}}\mathbf{Q}^{-\frac{1}{2}})^{-1}\mathbf{Q}^{-\frac{1}{2}}(l_{p,j}\mathbf{V}_2^*\tilde{\mathbf{S}}_2\mathbf{V}_1 - \mathbf{D}_2^{\frac{1}{2}}\mathbf{U}_2^*\tilde{\mathbf{S}}_1\mathbf{U}_1\mathbf{D}_1^{\frac{1}{2}}) \Big| \\ = & \Big| \frac{l_{p,j}}{n_2}\mathbf{V}_1^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_1 - \frac{l_{p,j}^2}{n_2^2}\mathbf{V}_1^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_2\mathbf{Q}^{-\frac{1}{2}}(l_{p,j}\mathbf{I} - \tilde{\mathbf{F}})^{-1}\mathbf{Q}^{-\frac{1}{2}}\mathbf{V}_2^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_1 \end{aligned}$$

$$
\begin{aligned}
&- \frac{l_{p,j}}{n_1}\mathbf{D}_1^{\frac{1}{2}}\mathbf{U}_1^*\mathbf{X}\big(l_{p,j}\mathbf{I}-\underline{\tilde{\mathbf{F}}}\big)^{-1}\mathbf{X}^*\mathbf{U}_1\mathbf{D}_1^{\frac{1}{2}}\\
&+ \frac{l_{p,j}}{n_2}\mathbf{V}_1^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_2\mathbf{Q}^{-\frac{1}{2}}\big(l_{p,j}\mathbf{I}-\tilde{\mathbf{F}}\big)^{-1}\mathbf{Q}^{-\frac{1}{2}}\frac{1}{n_1}\mathbf{D}_2^{\frac{1}{2}}\mathbf{U}_2^*\mathbf{X}\mathbf{X}^*\mathbf{U}_1\mathbf{D}_1^{\frac{1}{2}}\\
&+ \frac{l_{p,j}}{n_1}\mathbf{D}_1^{\frac{1}{2}}\mathbf{U}_1^*\mathbf{X}\mathbf{X}^*\mathbf{U}_2\mathbf{D}_2^{\frac{1}{2}}\mathbf{Q}^{-\frac{1}{2}}(l_{p,j}\mathbf{I}-\tilde{\mathbf{F}})^{-1}\mathbf{Q}^{-\frac{1}{2}}\frac{1}{n_2}\mathbf{V}_2^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_1\Big|\\
&= \Big|\{\psi_{n,k}+c_2\psi_{n,k}^2 m(\psi_{n,k})\}\mathbf{I}_M+\psi_{n,k}\underline{m}(\psi_{n,k})\mathbf{D}_1+\frac{1}{\sqrt{p}}\gamma_{kj}\psi_{n,k}\mathbf{I}_M\\
&\quad + \mathbf{B}_1(l_{p,j})+\mathbf{B}_2(l_{p,j})+\frac{\psi_{n,k}}{\sqrt{p}}\boldsymbol{\Omega}_M(\psi_{n,k},\mathbf{X},\mathbf{Y})+o(\frac{\psi_{n,k}}{\sqrt{p}})\Big|, \qquad (2.10)
\end{aligned}
$$

where $\mathbf{Q}=\mathbf{V}_2^*\tilde{\mathbf{S}}_2\mathbf{V}_2$, $\tilde{\mathbf{F}}$ and $\underline{\tilde{\mathbf{F}}}$ are defined as

$$
\tilde{\mathbf{F}}=\frac{1}{n_1}\mathbf{Q}^{-\frac{1}{2}}\mathbf{D}_2^{\frac{1}{2}}\mathbf{U}_2^*\mathbf{X}\mathbf{X}^*\mathbf{U}_2\mathbf{D}_2^{\frac{1}{2}}\mathbf{Q}^{-\frac{1}{2}},\ \underline{\tilde{\mathbf{F}}}=\frac{1}{n_1}\mathbf{X}^*\mathbf{U}_2\mathbf{D}_2^{\frac{1}{2}}\mathbf{Q}^{-1}\mathbf{D}_2^{\frac{1}{2}}\mathbf{U}_2^*\mathbf{X}. \qquad (2.11)
$$

Let $m(\lambda)$ and $\underline{m}(\lambda)$ be the Stieltjes transforms of the limiting spectral distributions of $\tilde{\mathbf{F}}$ and $\underline{\tilde{\mathbf{F}}}$, respectively. Furthermore, instead of ψ_k, $\psi_{n,k}$ is used to avoid the slow convergence as mentioned in Section 2.1, $\gamma_{kj}=\sqrt{p}\big(l_{p,j}/\psi_{n,k}-1\big)$, $j\in\mathcal{J}_k$, and

$$
\begin{aligned}
\mathbf{B}_1(l_{p,j}) =& \frac{\psi_{n,k}^2}{n_2^2}\mathbf{V}_1^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_2\mathbf{Q}^{-\frac{1}{2}}(\psi_{n,k}\mathbf{I}-\tilde{\mathbf{F}})^{-1}\mathbf{Q}^{-\frac{1}{2}}\mathbf{V}_2^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_1\\
&- \frac{l_{p,j}^2}{n_2^2}\mathbf{V}_1^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_2\mathbf{Q}^{-\frac{1}{2}}(l_{p,j}\mathbf{I}-\tilde{\mathbf{F}})^{-1}\mathbf{Q}^{-\frac{1}{2}}\mathbf{V}_2^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_1;\\
\mathbf{B}_2(l_{p,j}) =& \frac{\psi_{n,k}}{n_1}\mathbf{D}_1^{\frac{1}{2}}\mathbf{U}_1^*\mathbf{X}\big(\psi_{n,k}\mathbf{I}-\underline{\tilde{\mathbf{F}}}\big)^{-1}\mathbf{X}^*\mathbf{U}_1\mathbf{D}_1^{\frac{1}{2}}-\frac{l_{p,j}}{n_1}\mathbf{D}_1^{\frac{1}{2}}\mathbf{U}_1^*\mathbf{X}\big(l_{p,j}\mathbf{I}-\underline{\tilde{\mathbf{F}}}\big)^{-1}\mathbf{X}^*\mathbf{U}_1\mathbf{D}_1^{\frac{1}{2}}.
\end{aligned}
$$

Moreover, the $\boldsymbol{\Omega}_M(\lambda,\mathbf{X},\mathbf{Y})$ is defined as

$$
\boldsymbol{\Omega}_M(\lambda,\mathbf{X},\mathbf{Y})=\sum_{j=1}^{5}\boldsymbol{\Omega}_{M,j}(\lambda,\mathbf{X},\mathbf{Y}), \qquad (2.12)
$$

where

$$
\begin{aligned}
\mathbf{\Omega}_{M,1}(\lambda,\mathbf{X},\mathbf{Y}) &= \sqrt{p}\mathbf{V}_1^*(\tilde{\mathbf{S}}_2-\mathbf{I}_p)\mathbf{V}_1,\\
\mathbf{\Omega}_{M,2}(\lambda,\mathbf{X},\mathbf{Y}) &= \frac{\sqrt{p}\lambda}{n_2}\Big\{\mathrm{tr}(\lambda\mathbf{I}-\tilde{\mathbf{F}})^{-1}\mathbf{I}-\frac{1}{n_2}\mathbf{V}_1^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_2\mathbf{Q}^{-\frac{1}{2}}(\lambda\mathbf{I}-\tilde{\mathbf{F}})^{-1}\mathbf{Q}^{-\frac{1}{2}}\mathbf{V}_2^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_1\Big\},\\
\mathbf{\Omega}_{M,3}(\lambda,\mathbf{X},\mathbf{Y}) &= \frac{\sqrt{p}}{\sqrt{n_1}\lambda}\mathbf{D}_1^{\frac{1}{2}}\Big[\frac{\lambda}{\sqrt{n_1}}\big\{\mathrm{tr}(\lambda\mathbf{I}-\underline{\tilde{\mathbf{F}}})^{-1}\mathbf{I}-\mathbf{U}_1^*\mathbf{X}\big(\lambda\mathbf{I}-\underline{\tilde{\mathbf{F}}}\big)^{-1}\mathbf{X}^*\mathbf{U}_1\big\}\Big]\mathbf{D}_1^{\frac{1}{2}},\\
\mathbf{\Omega}_{M,4}(\lambda,\mathbf{X},\mathbf{Y}) &= \frac{\sqrt{p}}{n_1n_2}\mathbf{V}_1^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_2\mathbf{Q}^{-\frac{1}{2}}\big(\lambda\mathbf{I}-\tilde{\mathbf{F}}\big)^{-1}\mathbf{Q}^{-\frac{1}{2}}\mathbf{D}_2^{\frac{1}{2}}\mathbf{U}_2^*\mathbf{X}\mathbf{X}^*\mathbf{U}_1\mathbf{D}_1^{\frac{1}{2}},\\
\mathbf{\Omega}_{M,5}(\lambda,\mathbf{X},\mathbf{Y}) &= \frac{\sqrt{p}}{n_1n_2}\mathbf{D}_1^{\frac{1}{2}}\mathbf{U}_1^*\mathbf{X}\mathbf{X}^*\mathbf{U}_2\mathbf{D}_2^{\frac{1}{2}}\mathbf{Q}^{-\frac{1}{2}}\big(\lambda\mathbf{I}-\tilde{\mathbf{F}}\big)^{-1}\mathbf{Q}^{-\frac{1}{2}}\mathbf{V}_2^*\mathbf{Y}\mathbf{Y}^*\mathbf{V}_1.
\end{aligned}
$$

Note that the covariance matrix between $\mathbf{U}_1^*\mathbf{X}$ and $\mathbf{V}_1^*\mathbf{Y}$ is a zero matrix $\mathbf{0}_{M\times M}$, then according to Lemma 2.7 in Bai and Silverstein (1998) and equation (9) in Jiang et al. (2021)), we also obtain that ψ_k satisfies the following equation

$$\psi_k + c_2\psi_k^2 m(\psi_k) + \psi_k\underline{m}(\psi_k)\alpha_k = 0. \tag{2.13}$$

So far, we find the fluctuation of $l_{p,j}, j\in\mathcal{J}_k$ is reflected by γ_{kj} and is related to the limiting properties of $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{\Omega}_M(\psi_{n,k},\mathbf{X},\mathbf{Y})$ in (2.10). To investigate the limiting properties of these matrices, we first establish the invariance principle of the generalized spiked Fisher matrix in the following theorem. The invariance principle implies that the limiting distribution of spiked eigenvalues from a generalized spiked Fisher matrix remains the same, provided that the population distributions satisfy Assumptions 1–5.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

Theorem 1. *(*INVARIANCE PRINCIPLE THEOREM*) Assuming that* $(\mathbf{X}, \mathbf{Y})$ *and* $(\mathbf{W}, \mathbf{Z})$ *are two pairs of double arrays, each of which satisfies Assumptions 1–5, then* $\mathbf{\Omega}_M(\lambda, \mathbf{X}, \mathbf{Y})$ *and* $\mathbf{\Omega}_M(\lambda, \mathbf{W}, \mathbf{Z})$ *have the same limiting distribution, provided that one of them has a limiting distribution.*

Due to limited space, the proof of Theorem 1 is given in the Supplement Material. According to Theorem 1, we may assume that $\mathbf{X}$ and $\mathbf{Y}$ consist of entries with i.i.d. standard normal variables when deriving the limiting distributions of $\mathbf{B}_1(l_{p,j})$, $\mathbf{B}_2(l_{p,j})$ and $\mathbf{\Omega}_M(\psi_{n,k}, \mathbf{X}, \mathbf{Y})$. Firstly, define $m_2(\lambda) = \int 1/(\lambda - x)^2 \mathrm{d}\tilde{F}(x)$, $\underline{m}_2(\lambda) = \int 1/(\lambda - x)^2 \mathrm{d}\underline{\tilde{F}}(x)$, where $\tilde{F}(x)$ and $\underline{\tilde{F}}(x)$ are the limiting spectral distributions of the matrices $\tilde{\mathbf{F}}$ and $\underline{\tilde{\mathbf{F}}}$, respectively. Then, we obtain that

$$\mathbf{B}_1(l_{p,j}) = \frac{1}{\sqrt{p}}\gamma_{kj}\Big\{c_2\psi_{n,k}^3 m_2(\psi_{n,k}) + 2c_2\psi_{n,k}^2 m(\psi_{n,k})\Big\}\mathbf{I}_M + o(\frac{\psi_{n,k}}{\sqrt{p}});$$
$$\mathbf{B}_2(l_{p,j}) = \frac{1}{\sqrt{p}}\gamma_{kj}\Big\{\psi_{n,k}^2 \underline{m}_2(\psi_{n,k}) + \psi_{n,k}\underline{m}(\psi_{n,k})\Big\}\mathbf{D}_1 + o(\frac{\psi_{n,k}}{\sqrt{p}}).$$

The details of the derivation are listed in the Supplement Material. Thus, it follows from equation (2.10) that

$$\begin{aligned} 0 = \bigg| &\{\psi_{n,k} + c_2\psi_{n,k}^2 m(\psi_{n,k})\}\mathbf{I}_M + \psi_{n,k}\underline{m}(\psi_{n,k})\mathbf{D}_1 + \frac{\psi_{n,k}}{\sqrt{p}}\mathbf{\Omega}_M(\psi_{n,k}, \mathbf{X}, \mathbf{Y}) \\ &+ \frac{1}{\sqrt{p}}\gamma_{kj}\Big[\Big\{\psi_{n,k} + c_2\psi_{n,k}^3 m_2(\psi_{n,k}) + 2c_2\psi_{n,k}^2 m(\psi_{n,k})\Big\}\mathbf{I}_M \\ &+ \Big\{\psi_{n,k}^2 \underline{m}_2(\psi_{n,k}) + \psi_{n,k}\underline{m}(\psi_{n,k})\Big\}\mathbf{D}_1\Big] + o(\frac{\psi_{n,k}}{\sqrt{p}})\bigg|. \end{aligned} \tag{2.14}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

Furthermore, the limiting distribution of $\mathbf{\Omega}_M(\psi_{n,k}, \mathbf{X}, \mathbf{Y})$ is derived in the following corollary by replacing the entries in $\mathbf{X}$ and $\mathbf{Y}$ with the i.i.d. standard normal variables. The detailed proof is deferred to the Supplement Material.

Corollary 1. *Suppose that* $\mathbf{X}$ *and* $\mathbf{Y}$ *satisfy Assumptions 1–5, and let*

$$\theta_k = c_2 + c_2^2\psi_k^2 m_2(\psi_k) + 2c_2^2\psi_k m(\psi_k) + c_1\alpha_k^2 \underline{m}_2(\psi_k) + 2c_1c_2\alpha_k m_3(\psi_k), \quad (2.15)$$

where $m_3(\lambda) = \int x/(\lambda - x)^2 d\tilde{F}(x)$. *Then, it holds that* $\mathbf{\Omega}_M(\psi_{n,k}, \mathbf{X}, \mathbf{Y})$ *converges weakly to an* $M \times M$ *Hermitian matrix* $\mathbf{\Omega}_{\psi_k}$, *where* $[\mathbf{\Omega}_{\psi_k}]_{kk}$ *stands for the k-th block of* $\mathbf{\Omega}_{\psi_k}$, *and* ${\theta_k}^{-1/2}[\mathbf{\Omega}_{\psi_k}]_{kk}$ *is Gaussian Orthogonal Ensemble (GOE) for the real case*[†]. *For the complex case, the* ${\theta_k}^{-1/2}[\mathbf{\Omega}_{\psi_k}]_{kk}$ *is Gaussian Unitary Ensemble (GUE)*[‡].

Remark 1. Assumption 5 may be replaced by the following assumption.

Assumption 5′. Denote $\mathbf{U}_1 = (u_{ts})$, and $\mathbf{V}_1 = (v_{ts})$.

$$\beta_{x,i_1j_1i_2j_2} = \lim_{n_1\to\infty} \sum_{t=1}^{p} \bar{u}_{ti_1}u_{tj_1}u_{ti_2}\bar{u}_{tj_2}\big[\mathrm{E}\{|x_{11}|^4\delta(|x_{11}| \le \sqrt{n_1})\} - 2 - q\big] < \infty,$$

[†]GOE: a symmetric matrix whose upper triangular entries are independent real Gaussian variables with the diagonal entries being i.i.d. $\mathcal{N}(0, 2)$ and the upper off-diagonal entries being i.i.d. $\mathcal{N}(0, 1)$.

[‡]GUE: the diagonal entries are i.i.d. real $\mathcal{N}(0, 1)$, and the off diagonal entries are i.i.d. complex $\mathcal{CN}(0, 1)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

$$\beta_{y,i_1j_1i_2j_2} = \lim_{n_2\to\infty}\sum_{t=1}^{p}\bar{v}_{ti_1}v_{tj_1}v_{ti_2}\bar{v}_{tj_2}\big[\mathrm{E}\{|y_{11}|^4\delta(|y_{11}|\leq\sqrt{n_2})\}-2-q\big]<\infty,$$

where $q = 1$ for real case and $q = 0$ for complex and $\delta(\cdot)$ is the indicator function.

Let $\mathbf{u}_i = (u_{1i}, \ldots, u_{pi})'$ be the i-th column of the matrix $\mathbf{U}_1$, and let $\mathbf{v}_j = (v_{1j}, \ldots, v_{pj})'$ be the j-th column of the matrix $\mathbf{V}_1$. Then, all the conclusions of Corollary 1 still hold, but the limiting distribution of $\mathbf{\Omega}_M(\psi_{n,k}, \mathbf{X}, \mathbf{Y})$ turns to an $M \times M$ Hermitian matrix $\mathbf{\Omega}_{\psi_k} = (\omega_{ij})$, which has Gaussian entries of mean zero and variance

$$\mathrm{Cov}(\omega_{i_1,j_1}, \omega_{i_2,j_2}) = \begin{cases} (q+1)\theta_k + \beta_{x,iiii}\nu_1 + \beta_{y,iiii}\nu_2, & i_1 = j_1 = i_2 = j_2 = i; \\ \theta_k + \beta_{x,ijij}\nu_1 + \beta_{y,ijij}\nu_2, & i_1 = i_2 = i \neq j_1 = j_2 = j; \\ \beta_{x,i_1j_1i_2j_2}\nu_1 + \beta_{y,i_1j_1i_2j_2}\nu_2, & \text{other cases.} \end{cases}$$

Here θ_k is defined in (2.15), $\nu_1 = c_1\alpha_k^2\underline{m}^2(\psi_k)$ and $\nu_2 = c_2\big\{1+c_2\psi_k m(\psi_k)\big\}^2$.

The proof of this remark is based on Corollary 1 and given in Supplement Material. This remark is used in the simulations of Case I under nonGaussian assumptions.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

3. CLT for generalized spiked Fisher matrices

As an application of the invariance principle, we apply it for establishing the CLT for the spiked eigenvalues of a generalized spiked Fisher matrix $\mathbf{F}$. As mentioned in Proposition 1, a set of m_k consecutive sample eigenvalues $\{l_{p,j}(\mathbf{F}), j \in \mathcal{J}_k\}$ converges to a limit ρ_k laying outside the support of the limiting spectral distribution (LSD) of $\mathbf{F}$. To improve upon the work in Wang and Yao (2017), we consider a more general spiked Fisher matrix, $\mathbf{F}$ in (2.3) and the renormalized random vector

$$\gamma_k = (\gamma_{kj}, j \in \mathcal{J}_k) := \left\{ \sqrt{p} \left[\frac{l_{p,j}(\mathbf{F})}{\psi_n(\alpha_k)} - 1 \right], j \in \mathcal{J}_k \right\}. \tag{3.1}$$

Let

$$\phi_k = 1 + c_2 \psi_k^2 m_2(\psi_k) + 2c_2 \psi_k m(\psi_k) + \alpha_k \psi_k \underline{m}_2(\psi_k) + \alpha_k \underline{m}(\psi_k). \tag{3.2}$$

Then the CLT for the renormalized random vector γ_k is provided in the following theorem.

Theorem 2. *Suppose that Assumptions 1–5 hold. For each distinct spiked eigenvalue α_k*[†] *with multiplicity m_k, the m_k-dimensional real vector γ_k defined in (3.1) converges weakly to the joint distribution of the m_k eigenvalues of Gaussian random matrix $-\left[\mathbf{\Omega}_{\psi_k}\right]_{kk}/\phi_k$. Furthermore, $\mathbf{\Omega}_{\psi_k}$ is defined in*

[†] $\psi'(\alpha_k) > 0$, Jiang et al. (2021)

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

Corollary 1 and $[\mathbf{\Omega}_{\psi_k}]_{kk}$ *is the* k*-th diagonal block of* $\mathbf{\Omega}_{\psi_k}$ *corresponding to the indices* $\{i, j \in \mathcal{J}_k\}$.

The proof of Theorem 2 is postponed to the Supplement Material.

Remark 2. Suppose that $\mathbf{X}, \mathbf{Y}$ satisfy Assumptions 1–4, but Assumption 5 is weakened to Assumption $5'$ in Remark 1. Then all the conclusions of Theorem 2 still hold, but the limiting distribution of $\mathbf{\Omega}_M(\psi_{n,k}, \mathbf{X}, \mathbf{Y})$ tends to an $M \times M$ Hermitian Gaussian matrix $\mathbf{\Omega}_{\phi_k} = (\omega_{st})$ whose variances and covariances are defined in Remark 1.

4. Applications

In this section, we present two applications of Theorem 2 in the linear regression model and statistical signal processing. The first application is to analyze the local power of the Roy Maximum Root test in linear regression model, while the second application is related to signal detection in wireless communication.

4.1 Linear regression model

Let us consider a p-dimensional linear regression model

$$\mathbf{w}_i = \mathbf{B}\mathbf{z}_i + \boldsymbol{\varepsilon}_i, i = 1, \ldots, n, \tag{4.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

where $\{\boldsymbol{\varepsilon}_i, i = 1, \cdots, n\}$ is a sequence of independent and identically distributed normal error vector $\mathcal{N}_p(0, \boldsymbol{\Sigma})$, $\mathbf{B}$ is a $p \times q_0$ regression matrix, and $(\mathbf{z}_i), i = 1, \cdots, n$ a sequence of known regression variables of dimension q_0. In this section, we assume that $n \geq p + q_0$ and the rank of $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)$ is q_0.

We define a block decomposition $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2)$ with q_1 and q_2 columns, respectively ($q_0 = q_1 + q_2$). We partition the regression variables $\{\mathbf{z}_i\}$ accordingly as in $\mathbf{z}_i = (\mathbf{z}_{i1}', \mathbf{z}_{i2}')'$. Our goal is to test the hypothesis that

$$\mathcal{H}_0 : \ \mathbf{B}_1 = \mathbf{B}_1^0 \quad \text{vs.} \quad \mathcal{H}_1 : \ \mathbf{B}_1 \neq \mathbf{B}_1^0 \ , \tag{4.2}$$

where $\mathbf{B}_1^0$ is a known matrix. Roy (1953) proposed λ_1, the maximum eigenvalue of $\mathbf{H}\mathbf{G}^{-1}$, to test on linear regression hypothesis (4.2), where

$$\mathbf{G} = n\hat{\boldsymbol{\Sigma}} = \sum_{i=1}^{n} (\mathbf{w}_i - \widehat{\mathbf{B}}\mathbf{z}_i)(\mathbf{w}_i - \widehat{\mathbf{B}}\mathbf{z}_i)'; \quad \mathbf{H} = (\widehat{\mathbf{B}}_1 - \mathbf{B}_1^0)\mathbf{A}_{11:2}(\widehat{\mathbf{B}}_1 - \mathbf{B}_1^0)',$$

where $\widehat{\mathbf{B}}$ is the maximum likelihood estimators of $\mathbf{B}$, $\widehat{\mathbf{B}}_1$ denotes as the former q_1 columns of $\widehat{\mathbf{B}}$, and $\mathbf{A}_{11:2} = \sum_{i=1}^n \mathbf{z}_{i1}\mathbf{z}_{i1}' - \sum_{i=1}^n \mathbf{z}_{i1}\mathbf{z}_{i2}' \left(\sum_{i=1}^n \mathbf{z}_{i2}\mathbf{z}_{i2}'\right)^{-1} \sum_{i=1}^n \mathbf{z}_{i2}\mathbf{z}_{i1}'$. The distribution of λ_1 can be obtained from the joint density by integrating over the supporting set of all eigenvalues. Roy (1945) developed a method of integration and provided the distribution of $\lambda_1/(1 + \lambda_1)$ when $p = 2$. However, this integration becomes more difficult with increasing dimensionality. By Lemmas 8.4.1 and 8.4.2 in Anderson (2003), it is known that

$\mathbf{G} \sim W_p(\mathbf{\Sigma}, n-q_0)$, $\mathbf{H} \sim W_p(\mathbf{\Sigma}, q_1)$, and they are independent of each other under the Gaussian assumption. Therefore, $(n-q_0)q_1^{-1}\mathbf{H}\mathbf{G}^{-1}$ can be viewed as a generalized spiked Fisher matrix. We consider the large-dimensional setting

$$\tilde{c}_{n_1} = p/q_1 \to \tilde{c}_1 \in (0, \infty), \ \tilde{c}_{n_2} = p/(n - q_0) \to \tilde{c}_2 \in (0, 1). \tag{4.3}$$

As shown in Han et al. (2016), the largest root of $(n - q_0)q_1^{-1}\mathbf{H}\mathbf{G}^{-1}$, denoted as $l_{p,1} = (n - q_0)q_1^{-1}\lambda_1$, follows the Tracy-Widom law under the null hypothesis. Its rejection region at the 0.05 significance level is determined as follows:

$$\{l_{p,1} > \psi_0 + \sigma_{tw} C_{0.95}\}, \tag{4.4}$$

where $\psi_0 = (1 + h)^2(1 - \tilde{c}_2)^{-2}$ is the limit of the largest root under the null hypothesis with $h^2 = \tilde{c}_1 + \tilde{c}_2 - \tilde{c}_1\tilde{c}_2$, and $C_{0.95}$ is the 95% percentile of the Tracy-Widom distribution.

The value of σ_{tw} is determined by solving several trigonometric equations as described in Han et al. (2016). In order to derive a simpler expression, Wang and Yao (2017) derived a result based on Han et al. (2016). However, we found that their result only depends on the dimensionality p and one of the sample sizes n_1. Upon comparing the results of the two approaches, we observed that they were not the same. Therefore, we recalculated the value of σ_{tw} following Han et al. (2016) and present the simplified expression below, which is identical to the one in Han et al. (2016), i.e.

$$\sigma_{tw}^3 = \frac{\tilde{c}_1^2(\tilde{c}_1 + h)^4(\tilde{c}_1 + \tilde{c}_2)^6}{(n - q_0 + q_1)^2 h \tilde{c}_2^2 \big\{(\tilde{c}_1 + \tilde{c}_2)^2 - \tilde{c}_2(\tilde{c}_1 + h)^2\big\}^4}.$$

Based on the rejection region (4.4), Theorem 2 is applied to derive the asymptotic distribution of $l_{p,1}$ and provide the power function under the alternative hypothesis. The details are as follows:

Theorem 3. *For testing hypothesis (4.2), if the large-dimensional limiting scheme (4.3) holds, then the asymptotic distribution of the largest sample eigenvalue of* $(n - q_0)q_1^{-1}\mathbf{H}\mathbf{G}^{-1}$ *is that*

$$\Lambda_1 = \sqrt{p}\Big(\frac{l_{p,1}}{\psi_{n,1}} - 1\Big)\Big/\sigma_1 \Rightarrow \mathcal{N}(0,1), \quad \text{under } \mathcal{H}_1, \tag{4.5}$$

where $\sigma_1^2 = 2\theta_1/\phi_k^2$ *for the general real case with Assumption 5 and* $\sigma_1^2 = (2\theta_1 + \beta_{x,iiii}\nu_1 + \beta_{y,iiii}\nu_2)/\phi_k^2$ *for the real case with Assumption* $5'$ *instead. Then, the power of Roy Maximum Root test on linear regression hypothesis is calculated by*

$$P_R(\psi_{n,1}) = \Phi\Big\{ - \frac{\sqrt{p}\big(\psi_0 + \sigma_{tw}C_{0.95} - \psi_{n,1}\big)}{\psi_{n,1}\sigma_1}\Big\}, \tag{4.6}$$

where $C_{0.95}$ *is 95% quantile of Tracy-Widom distribution and* Φ *is the cumulative distribution function (cdf) of the standard Gaussian distribution.*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

In practice, the parameters $\psi_{n,1}$ and σ_1 in the asymptotic distribution (4.5) are involved with the unknown population largest spike α_1 and the LSD of $(n-q_0)q_1^{-1}\mathbf{H}\mathbf{G}^{-1}$. Therefore, we provide some estimators to calculate the estimated test statistic $\hat{\Lambda}_1$ instead of Λ_1 in (4.5) as follows.

First, for the population spike α_1, it follows from the first equation in (2.13) that the equation $1+\tilde{c}_2 l_{p,1} m(l_{p,1})+\underline{m}(l_{p,1})\alpha_1 = 0$ holds approximately. Thus we get the estimation of α_1,

$$\hat{\alpha}_1 = -\frac{1+\tilde{c}_2 l_{p,1} m(l_{p,1})}{\underline{m}(l_{p,1})},$$

where $m(l_{p,1})$ and $\underline{m}(l_{p,1})$ can be respectively estimated by

$$\hat{m}(l_{p,1}) = \frac{1}{p-|\mathcal{J}_1|}\sum_{i\notin\mathcal{J}_1}(l_{p,i}-l_{p,1})^{-1} \quad \text{and} \quad \hat{\underline{m}}(l_{p,1}) = -\frac{1-\tilde{c}_1}{l_{p,1}} + \tilde{c}_1\hat{m}(l_{p,1}).$$

Here r_i and $\mathcal{J}_1$ are defined as $r_i = |l_{p,i}-l_{p,1}|/|l_{p,1}|$ and $\mathcal{J}_1 = \{i \in (1,\cdots,p) : r_i \le 0.2\}$. The set $\mathcal{J}_1$ is selected to avoid the effect of multiple roots and make the estimator more accurate. The constant 0.2 is a more suitable threshold value according to our simulated results. Moreover, the following estimators may be used to calculate $\psi_{n,1}$ and σ_1,

$$\begin{aligned}
\hat{m}(\psi_{n,1}) &= \frac{1}{p}\sum_{i\in\mathcal{J}_1}(l_{p,i}-\hat{\psi}_{n,1})^{-1};\ \hat{\underline{m}}(\psi_{n,1}) = -\frac{1-\tilde{c}_1}{\hat{\psi}_{n,1}} + \tilde{c}_1\hat{m}(\hat{\psi}_{n,1});\\
\hat{m}_2(\psi_{n,1}) &= \frac{1}{p}\sum_{i\in\mathcal{J}_1}(l_{p,i}-\hat{\psi}_{n,1})^{-2};\ \hat{\underline{m}}_2(\psi_{n,1}) = \frac{1-\tilde{c}_1}{(\hat{\psi}_{n,1})^2} + \tilde{c}_1\hat{m}_2(\hat{\psi}_{n,1});\\
\hat{m}_3(\psi_{n,1}) &= \frac{1}{p}\sum_{i\in\mathcal{J}_1} l_{p,i}(l_{p,i}-\hat{\psi}_{n,1})^{-2};\ \hat{\psi}_{n,1} = \psi(\hat{\alpha}_1).
\end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

Thus the Λ_1 in (4.5) can be estimated by using these estimators.

4.2 Signal detection

The literature Han et al. (2016) established the Tracy-Widom law for the largest eigenvalue of a Fisher matrix and applied the results to the signal detection problem. In signal detection or cognitive, the model generally has the following form:

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \mathbf{\Sigma}^{1/2}\boldsymbol{e}_t, \quad t = 1, 2, \cdots, m,$$

where $\mathbf{y}_t$ is a p-dimensional observation, $\mathbf{A}$ is a $p \times k$ mixing matrix, $\mathbf{x}_t$ is a $k \times 1$ low-dimensional signal with covariance matrix $\mathbf{I}_k$, while $\boldsymbol{e}_t$ is an i.i.d. noise with covariance matrix $\mathbf{I}_p$. The signal $\mathbf{x}_t$ is independent with the noise $\boldsymbol{e}_t$. For more details, see Zeng and Liang (2009); Nadakuditi and Silverstein (2010). A fundamental task in signal processing is to test

$$\mathcal{H}_0 : \ \mathbf{A} = \mathbf{0} \quad \text{vs.} \quad \mathcal{H}_\mathbf{1} : \ \mathbf{A} \neq \mathbf{0}. \tag{4.7}$$

In engineering, one can have additional independent noise-only observations $\mathbf{z}_t = \mathbf{\Sigma}^{1/2}\boldsymbol{e}_t, t = 1, \cdots, T$. Let

$$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m), \quad \mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T).$$

We define the Fisher matrix as

$$\mathbf{F} = \frac{T}{m}(\mathbf{Z}\mathbf{Z}^*)^{-1}(\mathbf{Y}\mathbf{Y}^*),$$

and use the symbols l_1 and β_1 to denote the largest eigenvalue of $\mathbf{F}$ and $\mathbf{\Sigma}^{-1}(\mathbf{A}\mathbf{A}^* + \mathbf{\Sigma})$, respectively. We use the statistic l_1 to test the hypothesis (4.7). According to Han et al. (2016), l_1 after scaling tends to follow the Tracy-Widom law under the null hypothesis $\mathbf{A} = \mathbf{0}$. Based on equation (4.5), we conclude the theoretical power for correlated noise detection as

$$P_R(\beta_1) = \Phi\Big[- \frac{\sqrt{p}\big\{\psi_0 + \sigma_{tw} C_{0.95} - \psi_{n,1}(\beta_1)\big\}}{\psi_{n,1}(\beta_1)\sigma_1} \Big], \tag{4.8}$$

where the notations defined similarly to Theorem 3. It is worth pointing out that Theorem 7.1 in Wang and Yao (2017) is a special case of (4.8).

5. Simulation Study

5.1 Simulations for Section 3

To verify the generality and performance of our proposed limiting results compared to Wang and Yao (2017), we conduct simulations under two scenarios. The first scenario assumes that $\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}$ is of diagonal block-wise form as assumed in Wang and Yao (2017). The second scenario allows a general spiked matrix for $\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}$ without diagonal assumption.

Case I: The matrix $\mathbf{T}_p^*\mathbf{T}_p$ is taken to be a finite-rank perturbation of an identity matrix $\mathbf{I}_p$, where $\mathbf{\Sigma}_2 = \mathbf{I}_p$ and $\mathbf{\Sigma}_1$ is an identity matrix with the spikes $(20, 0.2, 0.1)$ of the multiplicity $(1, 2, 1)$ in the descending

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

order and thus $K = 3$ and $M = 4$.

Case II: The matrix $\mathbf{T}_p^*\mathbf{T}_p$ is a general positive definite matrix, but it does not necessarily have a diagonal block-wise independence assumption as proposed in Wang and Yao (2017). Let $\mathbf{\Sigma}_2 = \mathbf{I}_p$ and $\mathbf{\Sigma}_1 = \mathbf{U}_0\mathbf{\Lambda}\mathbf{U}_0^*$, where $\mathbf{\Lambda}$ is a diagonal matrix consisting of the spikes $(20, 0.2, 0.1)$ with multiplicity $(1, 2, 1)$, and the other eigenvalues are 1 in the descending order. Let $\mathbf{U}_0$ be equal to the matrix composed of eigenvectors of the $p \times p$ matrix $(\rho^{|i-j|})_{i,j=1,\ldots,p}$ with $\rho = 0.5$.

For each scenario, we consider two populations as following: In the first population, x_{ij} and y_{ij} are both i.i.d. samples from $N(0, 1)$. In the second population, x_{ij} and y_{ij} are i.i.d. samples from $P\{x_{ij} = \pm 1\} = P\{y_{ij} = \pm 1\} = 1/2$. Thus, $\mathrm{E}|x_{ij}|^4 = \mathrm{E}|y_{ij}|^4 = 1$. This aims to illustrate the invariance principle of large-dimensional spiked Fisher matrices.

To further demonstrate the validity of the CLT derived in Section 3 for a distribution with infinite fourth moments under Assumption 5, we generate i.i.d. samples x_{ij} and y_{ij} from $2^{-1/2}t(4)$ population distribution under the setting of Case II. In this setting, $\mathrm{E}x_{ij} = \mathrm{E}y_{ij} = 0$, $\mathrm{E}x_{ij}^2 = \mathrm{E}y_{ij}^2 = 1$, while the fourth moments of x_{ij}, y_{ij} are infinite. Since Assumptions $5'$ for the CLT derived in Section 3 is not valid for the distribution $t(4)$ under the setting of Case I, then the new CLT does not hold for $t(4)$ with Case

I. Furthermore, the limiting distribution for the two-sample spiked model derived in Wang and Yao (2017) is not applicable for $t(4)$ either. Therefore, we only examine the performance of the newly derived limiting distribution. In this simulation, we set $p = 200$, $n_1 = 1000$ and $n_2 = 400$, and we conduct 1000 replications for each case.

For Case I, Remark 2 can be applied. For the largest and least simple population spikes $\alpha_1 = 20$ and $\alpha_3 = 0.1$, we obtain the following CLTs:

$$\gamma_k = \sqrt{p-4}\left\{\frac{l_{p,j}(\mathbf{F})}{\psi_{n,k}} - 1\right\} \to N(0, \sigma_k^2),$$

where $j = 1$ for $k = 1$ and $j = p$ for $k = 3$, $\psi_{n,1} = 42.667, \sigma_1^2 = 2.383$ and $\psi_{n,3} = 0.0737, \sigma_3^2 = 1.343$ under the Gaussian assumption; meanwhile, $\sigma_1^2 = 1.116$ and $\sigma_3^2 = 0.180$ for the distribution with binary outcome. To improve accuracy, we use $p - M$ instead of p in the calculation.

For the spikes $\alpha_2 = 0.2$ with multiplicity 2, we consider the sample eigenvalues $l_{p,p-1}$, $l_{p,p-2}$, and obtain that the two-dimensional random vector

$$\gamma_2 = (\gamma_{21}, \gamma_{22})' = \left[\sqrt{p-4}\Big\{\frac{l_{p,p-2}(\mathbf{F})}{\psi_{n,2}} - 1\Big\}, \sqrt{p-4}\Big\{\frac{l_{p,p-1}(\mathbf{F})}{\psi_{n,2}} - 1\Big\}\right]'$$

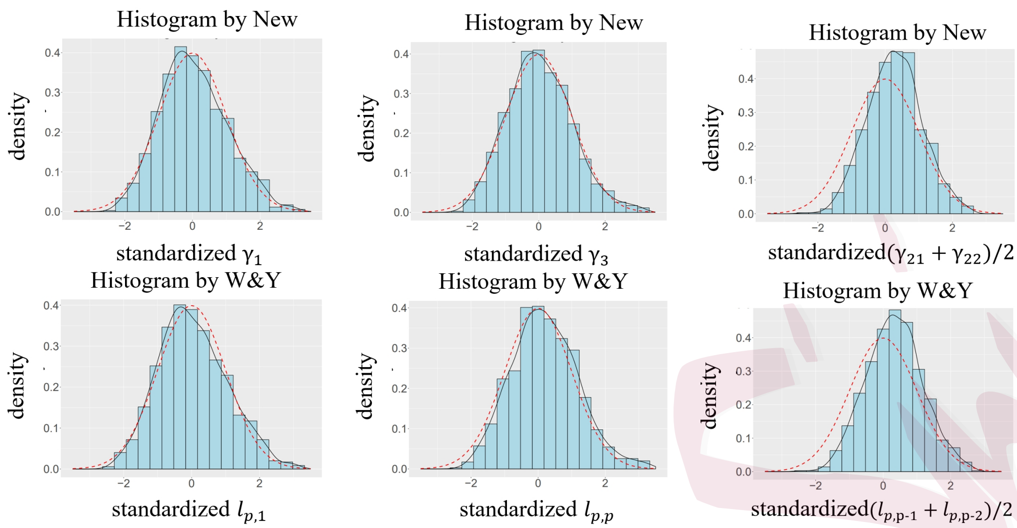
converges to the eigenvalues of random matrix $-\left[\mathbf{\Omega}_{\psi_2}\right]_{22}/\phi_k$, where $\psi_{n,2} =$ 0.133, $\phi_k = 1.439$ for the spike $\alpha_2 = 0.2$. Furthermore, the matrix $\left[\mathbf{\Omega}_{\psi_2}\right]_{22}$ is a 2×2 symmetric matrix with the independent Gaussian entries, of which the (s, t)-th element has mean zero and the variance given by $var(w_{st}) =$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

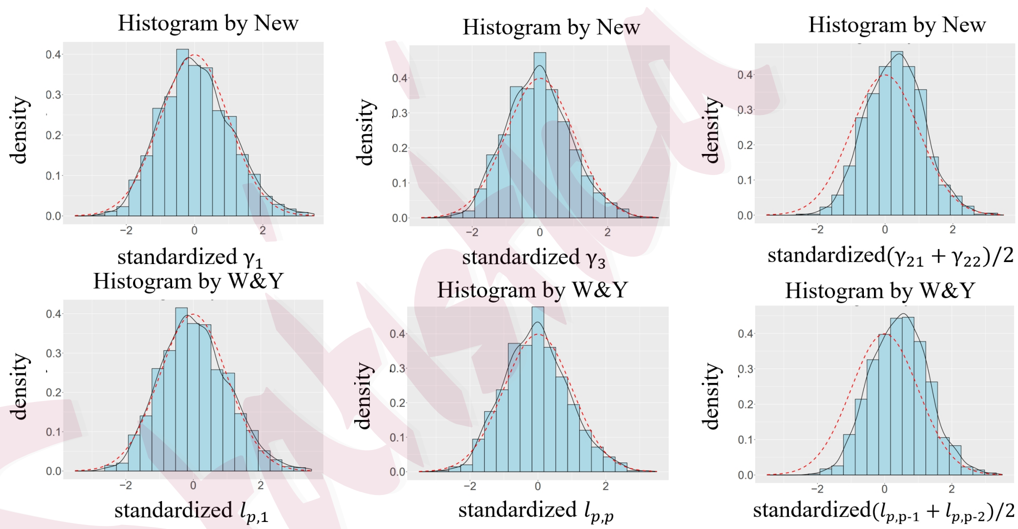
Table 1: KS Statistic and Percentiles of asymptotical distributions of the standardized γ_1, $\gamma_2^* = (\gamma_{21}+\gamma_{22})/2$, and γ_3 derived by our new method and Wang and Yao (2017).

E.V.	Case	Method	1%	5%	10%	25%	50%	75%	90%	95%	99%	KS
	Limiting	$N(0,1)$	-2.326	-1.645	-1.282	-0.674	0	0.674	1.282	1.645	2.326	–
		$x_{ij} \sim N(0,1)$ and $y_{ij} \sim N(0,1)$										
γ_1	I	New	-2.005	-1.455	-1.175	-0.650	-0.043	0.680	1.400	1.791	2.606	0.025
		WY	-2.047	-1.487	-1.200	-0.665	-0.045	0.694	1.429	1.828	2.661	0.028
γ_1	II	New	-1.996	-1.540	-1.191	-0.658	-0.009	0.671	1.378	1.775	2.660	0.031
		WY	-1.975	-1.524	-1.178	-0.652	-0.009	0.663	1.362	1.755	2.630	0.030
γ_2^*	I	New	-1.493	-1.017	-0.779	-0.257	0.265	0.814	1.330	1.631	2.233	0.150
		WY	-1.500	-1.005	-0.762	-0.226	0.313	0.880	1.412	1.718	2.351	0.162
γ_2^*	II	New	-1.574	-1.065	-0.761	-0.301	0.307	0.857	1.301	1.682	2.386	0.137
		WY	-1.456	-0.925	-0.624	-0.160	0.452	1.007	1.459	1.838	2.555	0.194
γ_3	I	New	-1.930	-1.502	-1.183	-0.609	-0.016	0.686	1.302	1.685	2.769	0.022
		WY	-1.802	-1.36	-1.038	-0.455	0.147	0.862	1.487	1.878	2.979	0.075
γ_3	II	New	-1.954	-1.554	-1.253	-0.703	-0.052	0.562	1.167	1.609	2.398	0.050
		WY	-1.963	-1.562	-1.259	-0.708	-0.055	0.561	1.167	1.611	2.405	0.051
		$P(x_{ij}=\pm 1)=1/2$ and $P(y_{ij}=\pm 1)=1/2$										
γ_1	I	New	-1.957	-1.518	-1.240	-0.646	-0.026	0.681	1.363	1.753	2.694	0.025
		WY	-1.968	-1.516	-1.152	-0.640	-0.028	0.694	1.391	1.788	2.749	0.026
γ_1	II	New	-2.019	-1.484	-1.187	-0.648	-0.007	0.637	1.410	1.823	2.503	0.023
		WY	-2.883	-2.119	-1.694	-0.925	-0.011	0.909	2.011	2.599	3.572	0.093
γ_2^*	I	New	-1.229	-0.812	-0.554	-0.029	-0.509	1.096	1.619	1.925	2.506	0.250
		WY	-0.378	-0.021	0.169	0.556	0.952	1.384	1.768	1.995	2.430	0.491
γ_2^*	II	New	-1.493	-0.956	-0.680	-0.178	0.403	0.918	1.476	1.798	2.438	0.185
		WY	-1.876	-1.132	-0.757	-0.072	0.721	1.417	2.180	2.629	3.496	0.267
γ_3	I	New	-2.381	-1.646	-1.296	-0.695	-0.013	0.578	1.114	1.520	2.217	0.047
		WY	-2.048	-1.272	-0.899	-0.268	-0.450	1.073	1.639	2.069	2.802	0.175
γ_3	II	New	-2.108	-1.527	-1.215	-0.666	0.017	0.711	1.357	1.733	2.615	0.026
		WY	-5.907	-4.283	-3.407	-1.870	0.056	1.982	3.789	4.842	7.308	0.236
		$x_{ij} \sim 2^{-1/2}t(4)$ and $y_{ij} \sim 2^{-1/2}t(4)$										
γ_1	II	New	-1.673	-1.312	-0.969	-0.418	0.253	0.939	1.703	2.199	3.324	0.112
γ_2^*	II	New	-3.936	-1.894	-1.343	-0.703	-0.029	0.517	1.004	1.324	1.811	0.068
γ_3	II	New	-2.873	-1.959	-1.512	-0.981	-0.274	0.444	1.112	1.510	2.249	0.118

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377



Case I



Case II

Figure 1: Histograms of standardized estimated eigenvalues over 1000 simulations when $x_{ij} \sim N(0,1)$ and $y_{ij} \sim N(0,1)$. The solid lines are the kernel density estimates, and the dashed lines are the probability density functions of $N(0,1)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

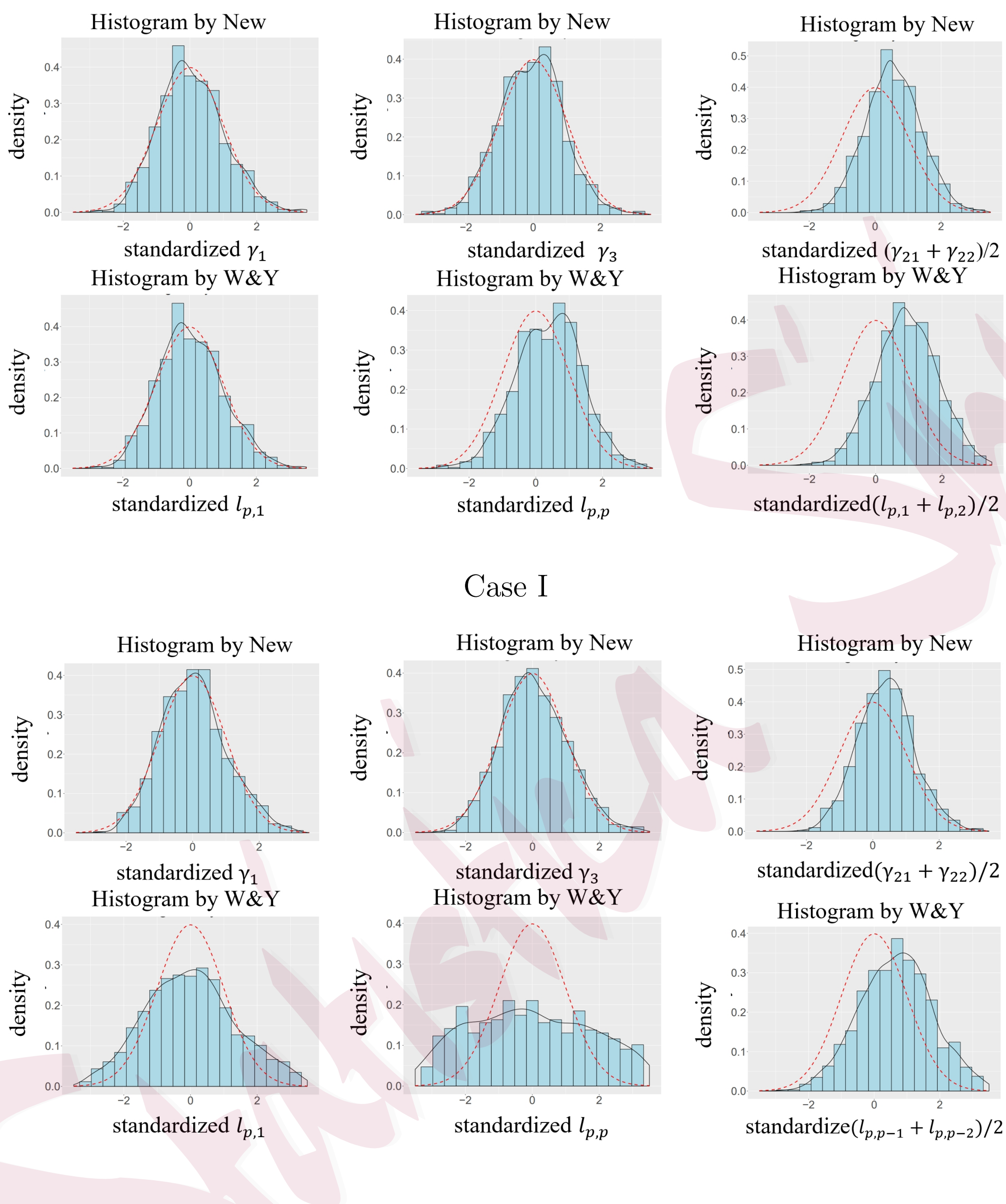


Figure 2: Histograms of standardized estimated eigenvalues over 1000 simulation when $P(x_{ij} = \pm 1) = P(y_{ij} = \pm 1) = 1/2$. The solid lines are the kernel density estimates, and the dashed lines are the probability density functions of $N(0, 1)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

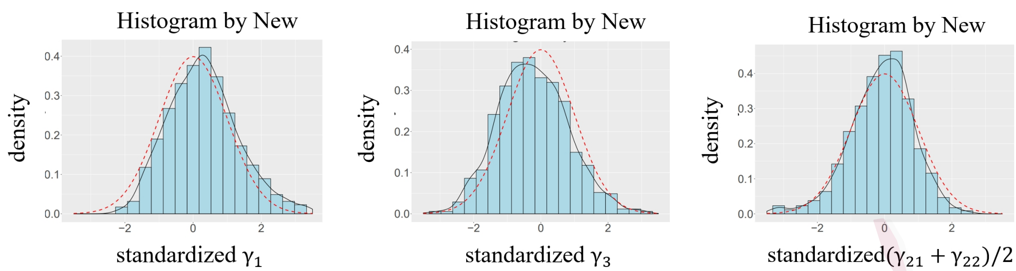


Figure 3: Histograms of standardized estimated eigenvalues over 1000 simulations under Case II when x_{ij} and y_{ij} follow the $2^{-1/2}t(4)$. The solid lines are the kernel density estimates, and the dashed lines are the probability density functions of $N(0,1)$.

1.163 if $s \neq t$ and $var(w_{st}) = 2.326$ if $s = t$ under the Gaussian assumption. All the results are the same except $var(w_{st}) = 0.502$ if $s = t$ under the second population (i.e. binary outcomes). Under Case II, it follows by Theorem 2 that the asymptotical means and covariances for all the three distributions considered in this simulation are the same as the ones of Case I under Gaussian assumption, even if the population fourth moments are infinite, such as $t(4)$.

Let $F(u)$ be the cdf of a random variable U, and $F_n(u)$ be its empirical cdf based on a sample $u_1, \cdots, u_n$. Define Kolmogorov-Smirnov (KS) statistic as follows:

$$KS = \sum_u |F_n(u) - F(u)|.$$

In our simulation, we set $F(u)$ to be the cdf of $N(0,1)$, and U be the standardized estimated eigenvalues of the generalized spiked Fisher matrix. We evaluate $F_n(u)$ based on 1000 simulations.

Table 1 depicts the nine typical percentiles of the empirical distributions of standardized estimated eigenvalues, The values of $\psi_{n,k}$ and σ_k^2 were calculated by two methods: one is derived in Theorem 2, and the other one is derived in Wang and Yao (2017), to compare the finite sample properties of our method with the one in Wang and Yao (2017). The corresponding KS statistics of the two methods are also reported in Table 1. To examine the overall pattern of the empirical distributions of the estimated spiked eigenvalues, Figures 1-3 show histograms based on 1000 simulations, along with the kernel density estimate and asymptotic limiting distributions.

Figure 1 depicts the histograms and kernel density estimates of the estimated eigenvalues when x_{ij} and y_{ij} follow $N(0,1)$. It can be observed that, under a Gaussian population, both the empirical distributions of our method and Wang and Yao's method are close to the asymptotical ones. This is further confirmed by the top panel of Table 1. The KS statistics of standardized γ_1 for Cases I and II and standardized γ_3 for Case II are very close for the two methods. Moreover, our newly proposed method has smaller KS statistics than Wang and Yao's method for the standardized γ_2^* in Cases I and II, as well as the standardized γ_3 in Case I.

Figure 2 depicts the histograms and kernel density estimates of the estimated eigenvalues when $P(x_{ij} = \pm 1) = P(y_{ij} = \pm 1) = 1/2$. Figure 2 clearly indicates that our method works well for both Case I and II, while Wang and Yao's method works well for Case I but not for Case II. The middle panel of Table 1 also delivers the same message. Except for standardized γ_1 in Case I, our method has much smaller KS statistics than Wang and Yao's method. For the percentiles, it seems that Wang and Yao's method has more spreading-out percentile. This implies it has larger variance.

Figure 3 depicts the histograms and kernel density estimates of the estimated eigenvalues when x_{ij} and y_{ij} follow $2^{-1/2}t(4)$. Figure 3 looks similar to the histograms for Case II in Figure 1. This implies that our method performs well when the fourth moment of population distribution is unbounded. This can be further confirmed by comparing the bottom panel of Table 1 and the corresponding ones in the top panel of Table 1.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

Table 2: Empirical sizes and powers

		$q_1/q_0 = 0.2$				$q_1/q_0 = 0.8$			
		Size		Power		Size		Power	
	p	New	CLRT	New	CLRT	New	CLRT	New	CLRT
$\tilde{c}_1 = 5, \tilde{c}_2 = 0.8;$	50	0.053	0.058	0.608	0.437	0.054	0.069	0.985	0.794
	100	0.043	0.053	0.974	0.933	0.051	0.046	1	0.899
	200	0.048	0.053	1	0.987	0.044	0.048	1	0.987
$\tilde{c}_1 = 5, \tilde{c}_2 = 0.5;$	50	0.043	0.057	1	0.998	0.051	0.049	1	1
	100	0.047	0.058	1	1	0.040	0.059	1	1
	200	0.042	0.048	1	1	0.045	0.051	1	1
$\tilde{c}_1 = 5, \tilde{c}_2 = 0.2;$	50	0.034	0.044	1	1	0.030	0.055	1	1
	100	0.038	0.049	1	1	0.043	0.052	1	1
	200	0.043	0.057	1	1	0.042	0.058	1	1
$\tilde{c}_1 = 2, \tilde{c}_2 = 0.8;$	50	0.042	0.056	0.547	0.467	0.051	0.058	0.838	0.598
	100	0.041	0.043	0.998	0.680	0.052	0.058	0.983	0.752
	200	0.038	0.055	1	0.847	0.051	0.049	0.979	0.885
$\tilde{c}_1 = 2, \tilde{c}_2 = 0.5;$	50	0.035	0.053	0.996	0.803	0.042	0.048	1	0.988
	100	0.052	0.056	1	0.995	0.041	0.052	1	1
	200	0.050	0.056	1	1	0.037	0.054	1	1
$\tilde{c}_1 = 2, \tilde{c}_2 = 0.2;$	50	0.041	0.062	1	1	0.041	0.046	1	1
	100	0.038	0.068	1	1	0.043	0.038	1	1
	200	0.047	0.046	1	1	0.038	0.049	1	1
$\tilde{c}_1 = 0.5, \tilde{c}_2 = 0.8;$	50	0.046	0.047	0.599	0.410	0.045	0.049	0.640	0.463
	100	0.055	0.049	0.734	0.569	0.058	0.054	0.695	0.504
	200	0.048	0.060	1	0.680	0.053	0.073	0.846	0.584
$\tilde{c}_1 = 0.5, \tilde{c}_2 = 0.5;$	50	0.047	0.048	1	0.699	0.045	0.043	0.985	0.785
	100	0.031	0.036	1	0.940	0.044	0.055	1	0.927
	200	0.051	0.061	1	1	0.049	0.057	1	0.978
$\tilde{c}_1 = 0.5, \tilde{c}_2 = 0.2;$	50	0.039	0.049	1	1	0.045	0.044	1	1
	100	0.037	0.047	1	1	0.044	0.053	1	1
	200	0.041	0.049	1	1	0.040	0.060	1	1

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0377

5.2 Numerical for Section 4

In this section, we conduct numerical study to compare the newly proposed test procedure for (4.2) in Section 4 with the corrected likelihood ratio test (CLRT) proposed by Bai et al. (2013). The simulation results for signal detections in Section 4.2 are similar to those for (4.2), we opt to omit them here to save space.

For testing hypothesis (4.2), we generate the elements of $\mathbf{B}_2$ from $\mathcal{N}(1,1)$ for each simulation. Under the null hypothesis, we set $\mathbf{B}_1 = \mathbf{0}$, while under the alternative hypothesis, half of the entries in the first column of $\mathbf{B}_1$ were generated from $\mathcal{N}(0.5,1)$ and the rest are zeros. We assume that the errors $\boldsymbol{\varepsilon}_i$ in (4.1) follows $\mathcal{N}_p(0,\mathbf{I}_p)$. All elements of $\mathbf{z}_i$ in the model are independent and identically distributed and are sampled from $\mathcal{N}(1,0.5)$.

We consider two cases: $q_1/q_0 = 0.8$ and $q_1/q_0 = 0.2$. For each case, set $p = 50, 100, 200$, $\tilde{c}_1 = 0.5, 2, 5$ and $\tilde{c}_2 = 0.2, 0.5, 0.8$. The limiting null distribution of Roy's test $\lambda_1(\mathbf{H}\mathbf{G}^{-1})$ follows the Tracy-Widom law for $\tilde{c}_1/\tilde{c}_2\lambda_1$ proposed by Han et al. (2016). In order to avoid the calculation of the complex integral in the Tracy-Widom law, we also derive the explicit expressions of the Tracy Widom law for $\tilde{c}_1/\tilde{c}_2\lambda_1$ by Theorem 1.12 in Bao et al. (2018) based on the functional relationship of canonical correlation matrix and Fisher matrix. Then we report both empirical sizes and powers with 1000 replications at a significance level $\alpha = 0.05$. The simulation results are summarized in the Tables 2.

The simulation illustrates that the limiting distribution of the Roy's test in linear regression model provides the slightly better sizes in our simulation settings. Table 2 shows that the proposed CLT for the Roy's test under the alternative hypothesis seems to be more powerful than the CLRT in this simulation setting. As seen from Table 2, the powers of the Roy's test rapidly increase to 1 as the sample size increases. For instance, for the case of $q_1/q_0 = 0.2, p = 50, \tilde{c}_1 = 2, \tilde{c}_2 = 0.8$ (i.e., $p = 50, n = 187, q_0 = 125, q_1 = 25$), the power is 0.547 and increases to 0.996 for the case of $q_1/q_0 = 0.2, p = 50, \tilde{c}_1 = 2, \tilde{c}_2 = 0.5$ (i.e., $p = 50, n = 225, q_0 = 125, q_1 = 25$). In general, both the Roy's test and the CLRT are expected to have good sizes, but our proposed approach has higher powers.

Supplementary Material

The supplementary material is available online and includes the additional proofs.

Acknowledgments

Dandan Jiang was supported by Key technologies for coordination and interoperation of power distribution service resource, Grant No. 2021YFB2401300; NSFC Grant No. 11971371 and Shaanxi Fundamental Science Research Project for Mathematics and Physics. Zhiqiang Hou's research was supported by NSFC Grant 12101359.

References

Anderson,T. W. (2003). An introduction to multivariate statistical analysis. 3rd ed. Wiley New York.

Baik, J., Arous, G. B. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, **33**, 1643–1697.

Bao, Z. G., Hu, J., Pan, G. M. and Zhou,W. (2018). Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *The Annals of Statistics*, **47**, 612–640.

Bai, Z. D., Jiang, D. D., Yao, J. F. and Zheng, S. R. (2013). Testing linear hypotheses in high-dimensional regressions. *Statistics: A Journal of Theoretical and Applied Statistics*, **47**(6), 1207–1223.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, **97**, 1382–1408.

Bai, Z. D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *The Annals of Prob-*

ability, **26**, 316–345.

Bai, Z. D. and Yao, J. F. (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, **44**, **(3)**, 447–474.

Bai, Z. D. and Yao, J. F. (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, **106**, 167–177.

Cai, T. T., Han, X. and Pan, G. M. (2020). Limiting laws for divergent spiked eigenvalues and largest non-spiked eigenvalue of sample covariance matrices. *The Annals of Statistics*, **48(3)**, 1255–1280.

Fan, J. and Wang, W. (2017). Asymptotics of empirical eigen-structure for ultra-high dimensional spiked covariance model. *The Annals of Statistics*, **45(3)**, 1342–1374.

Han, X., Pan, G. M. and Zhang, B. (2016). The Tracy-Widom Law for the largest eigenvalue of the F Type matrices. *The Annals of Statistics*, **44(4)**, 1564–1592.

Jiang, D. D. and Bai, Z. D. (2021). Generalized four moment theorem and an application to CLT for spiked eigenvalues of large-dimensional covariance matrices. *Bernoulli*, **27(1)**, 274–294.

Jiang, D. D, Hou, Z. Q. and Hu, J. (2021). The limits of the sample spiked eigenvalues for a high-dimensional generalized Fisher matrix and its applications. *Journal of Statistical Planning and Inference*, **3**, 208–217.

James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Annals of Mathematical Statistics*, **35**, 475–501.

Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, **29**, 295–327.

Johnstone, I. M. and Onatski, A. (2020). Testing in high-dimensional spiked models. *The Annals of Statistics*, **48(3)**, 1231–1254.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, **17**, 1617–1642.

Nadakuditi, R. R. and Silverstein, J. W. (2010). Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise only samples. *IEEE J. Sel. Top. Signal Process*, **4**, 468-480.

Roy, S. N. (1945). The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null-hypothesis. *Sankhyā*, **7**, 133–158.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Annals of Mathematical Statistics*, **24**, 220–238.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of the eigenvalues of large dimensional random matrices. *J. Multivariate Anal.*, **54**, 175–192.

Wang, Q. W. and Yao, J. F. (2017). Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *Annals of Statistics*, **45(1)**, 415–445.

Xie, J. S., Zeng, Y. C. and Zhu, L. X (2021). Limiting laws for extreme eigenvalues of large-dimensional spiked Fisher matrices with a divergent number of spikes. *Journal of Multivariate Analysis*, **184**, 104742.

Zeng, Y. H. and Liang, Y. C. (2009). Eigenvalue-based spectrum sensing algorithms for cognitive radio. *IEEE Transactions Communications* **57** 1784–1793.

Zheng, S. R., Bai, Z. D. and Yao, J. F. (2017). CLT for eigenvalue statistics of large dimensional general Fisher matrices with applications. *Bernoulli*, **23(2)**, 1130–1178.