

# ASYMPTOTIC JOINT DISTRIBUTION OF EXTREME EIGENVALUES AND TRACE OF LARGE SAMPLE COVARIANCE MATRIX IN A GENERALIZED SPIKED POPULATION MODEL

BY ZENG LI<sup>1</sup>, FANG HAN<sup>2</sup> AND JIANFENG YAO<sup>3</sup>

<sup>1</sup>Department of Statistics and Data Science, Southern University of Science and Technology, [liz9@sustech.edu.cn](mailto:liz9@sustech.edu.cn)

<sup>2</sup>Department of Statistics, University of Washington, [fanghan@uw.edu](mailto:fanghan@uw.edu)

<sup>3</sup>Department of Statistics and Actuarial Science, University of Hong Kong, [jeffyyao@hku.hk](mailto:jeffyyao@hku.hk)

This paper studies the joint limiting behavior of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model, where the asymptotic regime is such that the dimension and sample size grow proportionally. The form of the joint limiting distribution is applied to conduct Johnson–Graybill-type tests, a family of approaches testing for signals in a statistical model. For this, higher order correction is further made, helping alleviate the impact of finite-sample bias. The proof rests on determining the joint asymptotic behavior of two classes of spectral processes, corresponding to the extreme and linear spectral statistics, respectively.

**1. Introduction.** Considering a sequence of independent and identically distributed (i.i.d.)  $p$ -dimensional real-valued random vectors  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  with zero mean and population covariance matrix  $\Sigma_p$ , the corresponding sample covariance matrix is defined as

$$(1.1) \quad \mathbf{S}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top,$$

with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  denoting the eigenvalues of  $\mathbf{S}_n$ . It is statistically fundamental and important to study the distributions of the  $m$  largest eigenvalues  $\lambda_1, \dots, \lambda_m$  and the trace,  $\text{Tr}(\mathbf{S}_n) = \sum_{j=1}^p \lambda_j$ , of  $\mathbf{S}_n$  as  $p = p_n$  grows to infinity with  $n$ .

Indeed, each of them has led to a large volume of literature. For results on extreme sample eigenvalues, [16] first introduced the spiked population model as the nonnull case where all eigenvalues of  $\Sigma_p$  are unit except for a fixed small number of spikes, that is,

$$(1.2) \quad \text{Spec}(\Sigma_p) = \{\alpha_1, \dots, \alpha_m, \underbrace{1, \dots, 1}_{p-m}\}.$$

Here, we define  $\text{Spec}(\mathbf{A})$  to be the sets of eigenvalues of matrix  $\mathbf{A}$ . Under the “null” (i.e.,  $\Sigma_p$  is the identity matrix  $\mathbf{I}_p$ ), [16] established the Tracy–Widom law for the largest eigenvalue of real Wishart matrix  $\mathbf{S}_n$ . Following Johnstone’s development, many efforts have been put into quantifying the effect caused by spiked eigenvalues  $\{\alpha_k, 1 \leq k \leq m\}$  on  $m$  extreme sample ones  $\{\lambda_k, 1 \leq k \leq m\}$ . To name a few, under Johnstone’s spiked model settings, [6] thoroughly studied the almost sure limits of the extreme sample eigenvalues under the *Marčenko–Pastur regime* when  $p, n \rightarrow \infty$ ,  $p/n \rightarrow y \in (0, \infty)$ . They found that these limits are different when the corresponding population spiked eigenvalues are larger or smaller than critical values  $1 + \sqrt{y}$  and  $1 - \sqrt{y}$ . Similar phase transition phenomenon of largest sample eigenvalues was shown in [5] for complex Gaussian population. Paul [23] further proved that a phase transition of eigenvectors also occurs with Gaussian observations. Bai and Yao [2] followed the

Received September 2018; revised June 2019.

*MSC2020 subject classifications.* Primary 62E20, 62H15; secondary 15B52.

*Key words and phrases.* Generalized spiked model, asymptotic distribution, extreme eigenvalues, trace, large sample covariance matrix, random matrix theory.

set-up of [6] and established central limit theorems (CLTs) for the extreme sample eigenvalues associated with spikes outside the interval  $[1 - \sqrt{y}, 1 + \sqrt{y}]$  under general population distributions. Bai and Yao [3] extended the theory in [2] to a generalized spiked population model where the base population covariance matrix is arbitrary.

In contrast to extreme sample eigenvalues, many important statistics in multivariate analysis can be expressed as linear functionals of eigenvalues of some random matrices, namely, linear spectral statistics (LSS).  $\text{Tr}(\mathbf{S}_n)$  is one of the most important examples. Limiting behaviors of LSS has been intensively studied in the literature. One of the most widely used results is [4], which first established the asymptotic normality for LSS of sample covariance matrix  $\mathbf{S}_n$  under the Marčenko–Pastur regime with some moment restrictions on data entries. Further refinement and extensions can be found in numerous follow-up works. To name a few, [31] studied CLT for LSS of sample covariance matrix when the population mean vector is unknown. Chen and Pan [9] investigated the ultrahigh dimensional case when the dimension  $p$  is much larger than the sample size  $n$ . They further established the asymptotic normality for LSS as  $p/n \rightarrow \infty$  and  $n \rightarrow \infty$ . Zheng et al. [32] removed the fourth-order moment condition in [4] and incorporated it into the limiting parameters. Zheng [30] derived a CLT for LSS of large dimensional general Fisher matrices. The limiting distribution of  $\text{Tr}(\mathbf{S}_n)$  is derivable by implementing these results.

Despite the substantial advances in both directions, to our knowledge, little has been made on investigating the joint distribution of extreme sample eigenvalues and trace, which is equivalent to studying the asymptotic joint distribution of the largest and summation of sample eigenvalues. As will be seen soon, obtaining such a limiting distribution is fundamental in many applications, and is worth investigating in depth.

As a first contribution of this paper, we aim to study such a joint distribution. For this, we focus on Bai and Yao's generalized spiked model [3], which generalizes Johnstone's spiked model in [16]. Here, the population covariance matrix  $\Sigma_p$  has the structure

$$(1.3) \quad \Sigma_p = \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{p'} \end{pmatrix},$$

where  $\Lambda$  and  $\mathbf{V}_{p'}$  are of dimension  $m \times m$  and  $p' \times p'$  ( $p' = p - m$ ) and  $\Lambda$  is assumed fixed. The eigenvalues of  $\Lambda$  and  $\mathbf{V}_{p'}$  are  $\alpha_1 \geq \dots \geq \alpha_m > 0$  and  $\beta_{p',1} \geq \dots \geq \beta_{p',p'} \geq 0$ , respectively. The  $\alpha_j$ 's are larger than and well separated from  $\beta_{p',j}$ 's, thus named as *spiked eigenvalues*.

Under the generalized spiked model with  $p/n \rightarrow y \in (0, \infty)$ , we prove that the extreme eigenvalues and  $\text{Tr}(\mathbf{S}_n)$  are jointly asymptotically normal and asymptotically independent. The results are hence connected to the influential work of [11] and [14] on sum and maximum of i.i.d. and strongly mixing random variables. The conclusion holds as long as finite fourth-order moments exist, and in particular, requires no normality assumption. Our result is hence also connected to another interesting related work [12] where for sample covariance matrix  $\mathbf{S}_n$  with heavy-tailed entries, this asymptotic independence also holds.

Although  $\text{Tr}(\mathbf{S}_n)$  can be represented as the summation of all sample eigenvalues, in fact it is very difficult to quantify the correlation between extreme eigenvalues and the rest bulky ones, especially under the high dimensional settings without a Gaussian assumption. In facing this challenge, we make full use of the spiked model structure and carry out a block-decomposition analysis of spiked and nonspiked ones. The correlation between extreme eigenvalues and trace of each block of  $\mathbf{S}_n$  is analyzed separately based on the joint asymptotic behavior of two classes of spectral processes, corresponding to the extreme and linear spectral statistics, respectively. The idea of block-decomposition provides a novel perspective for proving the asymptotic independency between  $(\lambda_k)_{1 \leq k \leq m}$  and  $\text{Tr}(\mathbf{S}_n)$ . [23] adopted a similar

block-decomposition technique to represent sample eigenvalues as solutions to certain equations. However, [23] only considered the Gaussian case and focused on individual behavior of eigenvalues and eigenvectors when their population covariance is spiked with unit bulk eigenvalues.

The form of joint limiting distribution is then employed to conduct Johnson–Graybill-type tests [15], a family of approaches testing for signals in a statistical model based on the sample ratio  $\lambda_1/(\text{Tr}(\mathbf{S}_n)/p)$ . To name an application, this family of tests is important in modern signal processing applications, such as testing for the presence of signals in cognitive radio and nonparametric signal detection in array processing. For more details, we refer the readers to [8] and a comprehensive review paper [24]. Johnstone [16] proved that when  $\Sigma_p = \sigma^2 \mathbf{I}_p$ ,  $\lambda_1/\sigma^2$  converges to the Tracy–Widom (TW) distribution after appropriate centering and scaling. However, this ratio test statistic  $\lambda_1/(\text{Tr}(\mathbf{S}_n)/p)$  cannot be well approximated by the same asymptotic distribution. Finite sample adjustment of critical values for every  $(p, n)$  combinations was made in [21]. Deo [13] suggested an alternative variance correction which also improved the finite sample approximation. However, neither of them derived the asymptotic distribution of this studentized statistic. Furthermore, in the presence of  $m$  spikes, the limiting TW distribution of  $\lambda_{m+1}/\sigma^2$  has not been fully testified yet. Deo [13] performed some simulation experiments of testing multiple spikes based on the TW conjecture of  $\lambda_{m+1}/\sigma^2$ . They found that the test was uniformly undersized due to the downward bias of the TW approximation. Further analytic tools are needed to correct this bias, which is nevertheless nontrivial.

In this paper, we start from a different perspective by studying the behavior of  $\lambda_m/\sigma^2$  in the presence of  $m$  spikes. We use  $(p-m)^{-1} \sum_{j>m} \lambda_j$  instead of  $p^{-1} \text{Tr}(\mathbf{S}_n)$  as the surrogate for  $\sigma^2$ . Although the analytic tools are the same, the former enjoys better performance in finite sample cases. As a specific example, we formulate our null hypothesis as the spiked covariance model where

$$(1.4) \quad \Sigma_p = \sum_{i=1}^m a_i \mathbf{v}_i \mathbf{v}_i^\top + \sigma^2 \mathbf{I}_p,$$

and  $\mathbf{v}_i$ 's are orthonormal vectors. A typical example of such parametrization originates from the factor model where the  $p$ -dimensional data vector  $\mathbf{X}_t$  has a factor structure of the form

$$\mathbf{X}_t = \mathbf{A} \mathbf{F}_t + \mathbf{E}_t,$$

with  $\mathbf{E}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_p)$  independent of  $\mathbf{F}_t \sim \mathcal{N}(0, \mathbf{I}_m)$  and  $\mathbf{A}_{p \times m}$  is a deterministic matrix such that  $\mathbf{A} \mathbf{A}^\top$  has spectrum  $\sum_{i=1}^m a_i \mathbf{v}_i \mathbf{v}_i^\top$ . The limiting distribution of our test statistic,  $\lambda_m / \frac{1}{p-m} \sum_{j>m} \lambda_j$ , is then derived based on the asymptotic joint distribution of  $((\lambda_k)_{1 \leq k \leq m}, \text{Tr}(\mathbf{S}_n))^\top$ , and the corresponding test is implementable due to the developed theory. Our test targets at detection of signals above certain signal-to-noise ratio. Higher order corrections are further made to alleviate the impact of finite sample bias, which ensures satisfactory testing size and power.

It is worth mentioning here that this test is closely related to sphericity test (i.e., to test  $H_0 : m = 0$ , v.s.  $H_1 : m > 0$ ) discussed in [22] and [27]. In particular, they have nonzero power under the spiked alternative (i.e.,  $H_1 : m > 0$ ) even when the spikes are below the phase transition threshold. Admittedly, our test statistic can only detect distant spikes above the phase transition threshold. However, the sphericity test in [22] and [27] is only designed for testing existence of signals while our tests can be used to detect total number of spikes and the signal strength of the spikes being tested. In another related work, [10] formulated the observed data matrix  $\mathbf{X} \in \mathbb{R}^{n \times p}$  as the sum of a low-rank signal matrix  $\mathbf{B} \in \mathbb{R}^{n \times p}$  and a Gaussian noise matrix  $\mathbf{E} \in \mathbb{R}^{n \times p}$  and aimed at finding the rank of the deterministic signal

matrix  $\mathbf{B}$ . It is, however, very different from our model settings. In their model, signals are treated as a low-rank mean of the observed data matrix while in this paper we analyze spiked models when factors are embedded in a spiked population covariance structure.

Throughout the paper, we use bold Roman capital letters to represent matrices, for example,  $\mathbf{A}$ .  $\text{Tr}(\mathbf{A})$  and  $|\mathbf{A}|$  denote the trace and determinant of matrix  $\mathbf{A}$ . For matrix  $\mathbf{A}$ ,  $[\mathbf{A}]_{ij}$  denotes the  $(i, j)$ th entry of  $\mathbf{A}$ .  $\text{diag}(\alpha_1, \dots, \alpha_m)$  represents an  $m \times m$  diagonal matrix with diagonal entries  $\alpha_1, \dots, \alpha_m$ . Scalars are often in lowercase letters and random ones in capitals. Vectors follow bold italic style like  $\mathbf{v}_i$  and random vectors are in capitals like  $\mathbf{F}_i$ .  $\mathbb{N}$ ,  $\mathbb{R}$  and  $\mathbb{C}$  represent the sets of natural, real and complex numbers.  $\mathbf{1}(\cdot)$  stands for indicator function and  $\mathbf{T}$  stands for transpose of vectors or matrices. Let  $f : \mathbb{C} \rightarrow \mathbb{C}$  be a complex-valued function defined on the complex plane  $\mathbb{C}$ , then  $\oint_{\gamma} f(z) dz$  denotes the contour integral of  $f(z)$  on the Jordan curve  $\gamma$ . For any  $x \in \mathbb{R}$ ,  $\delta_x$  represents the point mass at  $x$ .

The remaining sections are organized as follows. Section 2 gives a detailed description of the generalized spiked model and introduces some preliminary results which form the basis of our analysis. Our main results are presented in Section 3. An application to factor modeling is studied in depth in Section 4. Proofs of theorems and technical lemmas are relegated to Section 5.

**2. Generalized spiked population model and preliminaries.** For any  $p \times p$  square matrix  $\mathbf{A}$  with eigenvalues  $(\theta_j)_{1 \leq j \leq p}$ , its empirical spectral distribution (ESD) is the measure  $F^{\mathbf{A}} = p^{-1} \sum_{j=1}^p \delta_{\theta_j}$  (weighting equally the eigenvalues). Under the generalized spiked population model (1.3), the following assumptions are made:

- (i) as  $n \rightarrow \infty$ ,  $p = p_n \rightarrow \infty$  such that  $p/n \rightarrow y \in (0, \infty)$ ;
- (ii) the sequence of spectral norms of  $\Sigma_p$  is bounded and the ESD  $H_{p'}$  of  $\mathbf{V}_{p'}$  converges to a nonrandom limiting distribution  $H$ ;
- (iii) the eigenvalues  $\{\beta_{p',j}, 1 \leq j \leq p'\}$  of  $\mathbf{V}_{p'}$  are such that as  $n \rightarrow \infty$ ,

$$\sup_{j \leq p'} d(\beta_{p',j}, \Gamma_H) = \varepsilon_p \rightarrow 0,$$

where  $d(x, A)$  denotes the Euclidean distance of  $x$  to a set  $A$  and  $\Gamma_H$  stands for the support of  $H$ ;

- (iv) the sample vectors  $\mathbf{X}_i, 1 \leq i \leq n$  can be expressed as  $\mathbf{X}_i = \Sigma_p^{1/2} \mathbf{Y}_i$ , where  $\mathbf{Y}_i$  are i.i.d.  $p$ -dimensional vectors with i.i.d. components  $\{Y_{ij}, j = 1, \dots, p\}$  satisfying  $\mathbb{E}Y_{ij} = 0, \mathbb{E}|Y_{ij}|^2 = 1$ , and  $\mathbb{E}|Y_{ij}|^4 = \nu_4 < \infty$ .

Letting  $\mu$  be a finite measure on the real line with support  $\Gamma_{\mu}$ , its Stieltjes transform  $s_{\mu}(z)$  is defined as

$$s_{\mu}(z) = \int \frac{1}{x - z} \mu(dx), \quad z \in \mathbb{C}^+,$$

where  $\mathbb{C}^+ := \{z \in \mathbb{C} : \Im(z) > 0\}$  is the upper half-plane with positive imaginary part and  $\Im(z)$  denotes the imaginary part of any given complex value  $z$ .

Let  $F^{\mathbf{S}_n}$  be the ESD of the sample covariance matrix  $\mathbf{S}_n$ . It is well known that under Assumptions (i) to (iv),  $F^{\mathbf{S}_n}$  weakly converges to a nonrandom probability measure  $F^{y,H}$ , the Marčenko–Pastur (M.P.) distribution with indexes  $(y, H)$ . Its Stieltjes transform  $s(z)$  is implicitly defined as a solution to the equation

$$s(z) = \int \frac{1}{t(1 - y - yzs(z)) - z} dH(t).$$

Correspondingly, the Stieltjes transform  $s_n(z) = \frac{1}{p} \text{Tr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1}$  of the ESD  $F^{\mathbf{S}_n}$  converges to  $s(z)$  almost surely as  $n \rightarrow \infty$ .

Moreover, consider an  $n \times n$  companion matrix for  $\mathbf{S}_n$ ,

$$\underline{\mathbf{S}}_n := \frac{1}{n} \mathbf{Y}^\top \Sigma_p \mathbf{Y} \quad \text{with } \mathbf{Y} = (Y_1, \dots, Y_n).$$

Both matrices share the same nonnull eigenvalues and their ESDs satisfy

$$nF^{\underline{\mathbf{S}}_n} - pF^{\mathbf{S}_n} = (n - p)\delta_0.$$

Their limits and respective Stieltjes transforms are linked to each other by the relation

$$\underline{F}^{y,H} - yF^{y,H} = (1 - y)\delta_0, \quad \underline{s}(z) = -\frac{1 - y}{z} + y s(z),$$

and the Stieltjes transform  $\underline{s}(z)$  of  $F^{\underline{\mathbf{S}}_n}$  satisfies the Silverstein equation [25]:

$$(2.1) \quad z = -\frac{1}{\underline{s}} + y \int \frac{t}{1 + t\underline{s}} dH(t).$$

Notice that the spiked structure (1.3) can be viewed as a finite rank perturbation of a general population covariance matrix with eigenvalues  $\{\beta_{p',j}\}$ . As the number of spikes  $m$  is fixed while  $p \rightarrow \infty$ , the limiting spectral distribution of  $F^{\mathbf{S}_n}$  is determined by the distribution of bulk population eigenvalues  $\{\beta_{p',j}\}$  independent of the spikes. However, the behavior of the  $m$  extreme sample eigenvalues  $\lambda_1, \dots, \lambda_m$  relies heavily on their population counterparts  $\alpha_1, \dots, \alpha_m$ .

Consider the functional inverse  $\psi$  of the function  $\alpha : x \mapsto -1/\underline{s}(x)$ . By (2.1), we have

$$(2.2) \quad \begin{aligned} \psi(\alpha) &= \psi_{y,H}(\alpha) = \alpha + y\alpha \int \frac{t}{\alpha - t} dH(t), \\ \psi'(\alpha) &= 1 - y \int \frac{t^2}{(\alpha - t)^2} dH(t), \quad \psi''(\alpha) = 2y \int \frac{t^2}{(\alpha - t)^3} dH(t). \end{aligned}$$

This function  $\psi(\cdot)$  is well defined for all  $\alpha \notin \Gamma_H$ .

Bai and Yao [3] gave a detailed characterization about the phase transition phenomenon of the limits of  $\lambda_1, \dots, \lambda_m$  when  $\alpha_1, \dots, \alpha_m$  satisfy different conditions. They name a generalized spiked eigenvalues  $\alpha$  a *distant spike* for the M.P distribution  $F^{y,H}$  if  $\psi'(\alpha) > 0$  and a *close spike* if  $\psi'(\alpha) \leq 0$ . Using the characterization of support of the LSD  $F^{y,H}$  given in [25], it can be seen that for distant spikes, the corresponding sample eigenvalues almost surely converge to limits which are outside the support  $\Gamma_{\underline{F}^{y,H}}$  of LSD of  $\mathbf{S}_n$ . These spikes are also referred as “outliers” in the literature.

In this paper, we are focused on the generalized spiked model with distant spikes. In addition, for presentation simplicity, we only consider the case when  $\alpha_1, \dots, \alpha_m$  are nonidentical. Extension to the case with possible overlaps on population spikes is straightforward using the developed techniques in this paper. Therefore, in addition to Assumptions (i) to (iv), we further assume that:

(v)  $\mathbf{\Lambda}$  is a fixed  $m \times m$  matrix with nonidentical bounded eigenvalues  $\alpha_1 > \dots > \alpha_m > \sup_p \max_j \beta_{p',j}$ . All  $\alpha'_k$ s are distant spiked eigenvalues satisfying  $\psi'(\alpha_k) > 0$ .

**3. Main results.** In this section, we study the asymptotic behavior of  $m$  largest sample eigenvalues  $\lambda_1, \dots, \lambda_m$  and trace of  $\mathbf{S}_n$ . Define the spectral decomposition of  $\mathbf{\Lambda} = (\Lambda_{ij})_{m \times m}$  to be

$$\mathbf{\Lambda} = \mathbf{U} \text{diag}(\alpha_1, \dots, \alpha_m) \mathbf{U}^\top,$$

where we remind that  $\text{diag}(\alpha_1, \dots, \alpha_m)$  represents the  $m \times m$  diagonal matrix with diagonal entries  $\alpha_1, \dots, \alpha_m$ .

**THEOREM 3.1.** *Under Assumptions (i) to (v), for the  $m$  largest eigenvalues  $\lambda_1, \dots, \lambda_m$  of  $\mathbf{S}_n$ , denoting  $\psi(\alpha_k)$  as  $\psi_k$ , we have, the  $m$ -dimensional random vector*

$$\left( \sqrt{n} \left( \frac{\lambda_1}{\psi_1} - 1 \right), \dots, \sqrt{n} \left( \frac{\lambda_m}{\psi_m} - 1 \right) \right)^\top$$

and  $\text{Tr}(\mathbf{S}_n)$  are jointly asymptotically normal and independent. Marginally,

$$\text{Tr}(\mathbf{S}_n) - \text{Tr}(\mathbf{\Sigma}_p) \xrightarrow{d} \mathcal{N}(0, 2y\gamma_2 + y(v_4 - 3)\gamma_{d,2}),$$

where for  $k = 1, 2, \dots$ ,  $\gamma_k := \int t^k dH(t)$  denotes the  $k$ th moment of LSD of  $\mathbf{V}_{p'}$  and  $\gamma_{d,2} := \lim_{p' \rightarrow \infty} \frac{1}{p'} \sum_{i=1}^{p'} [\mathbf{V}_{p'}]_{ii}^2$ .

Moreover,  $(\sqrt{n}(\frac{\lambda_1}{\psi_1} - 1), \dots, \sqrt{n}(\frac{\lambda_m}{\psi_m} - 1))^\top$  weakly converges to an  $m$ -dimensional Gaussian vector  $(M_1, \dots, M_m)^\top$ , with each

$$M_k = \mathbf{u}_k^\top \mathbf{G}(\psi(\alpha_k)) \mathbf{u}_k,$$

where  $\mathbf{u}_k = (u_{1k}, \dots, u_{mk})^\top$  is the  $k$ th column of  $\mathbf{U}$ ,  $\mathbf{G}(\psi(\alpha_k))$  is an  $m \times m$  Gaussian random matrix with independent entries such that

(a) its diagonal elements are i.i.d. Gaussian with mean zero and variance

$$(3.1) \quad \sigma_{\alpha_k}^2 := 2 \frac{\alpha_k^2}{\psi_k^2} \cdot \psi'(\alpha_k) + \beta_y \frac{\alpha_k^2}{\psi_k^2} \cdot (\psi'(\alpha_k))^2,$$

with  $\beta_y = \mathbb{E}|Y_{ij}|^4 - 3 = v_4 - 3$  denoting the fourth cumulant of base entries  $\{Y_{ij}\}$ ;

(b) its upper triangular elements are i.i.d. Gaussian, with mean zero and variance

$$(3.2) \quad s_{\alpha_k}^2 = \frac{\alpha_k}{\psi_k} \cdot \psi'(\alpha_k).$$

Meanwhile, denoting the  $(i, j)$ th entry of matrix  $\mathbf{A}$  by  $[\mathbf{A}]_{ij}$ , we have, for  $k_1 \neq k_2$ ,

$$\text{Cov}([\mathbf{G}(\psi(\alpha_{k_1}))]_{ij}, [\mathbf{G}(\psi(\alpha_{k_2}))]_{ij}) = \frac{\alpha_{k_1} \alpha_{k_2} \psi'(\alpha_{k_1}) \psi'(\alpha_{k_2})}{\psi_{k_1} \psi_{k_2}} \cdot \frac{\alpha_{k_1} - \alpha_{k_2}}{\psi_{k_1} - \psi_{k_2}}, \quad i \neq j,$$

$$\text{Cov}([\mathbf{G}(\psi(\alpha_{k_1}))]_{ii}, [\mathbf{G}(\psi(\alpha_{k_2}))]_{ii}) = \frac{\alpha_{k_1} \alpha_{k_2} \psi'(\alpha_{k_1}) \psi'(\alpha_{k_2})}{\psi_{k_1} \psi_{k_2}} \cdot \left( 2 \cdot \frac{\alpha_{k_1} - \alpha_{k_2}}{\psi_{k_1} - \psi_{k_2}} + \beta_y \right),$$

$$\text{Cov}([\mathbf{G}(\psi(\alpha_{k_1}))]_{ij}, [\mathbf{G}(\psi(\alpha_{k_2}))]_{i'j'}) = 0 \quad \text{for all other cases.}$$

Of note, [26] studied the limiting distribution of the random vector

$$\left( \sqrt{n} \left( \frac{\lambda_1}{\psi_1} - 1 \right), \dots, \sqrt{n} \left( \frac{\lambda_m}{\psi_m} - 1 \right) \right)^\top$$

under Johnstone’s spiked model (1.2) with  $\mathbf{V}_{p'} = \mathbf{I}_{p'}$ . Here, we allow a more general  $\mathbf{V}_{p'}$ .

Following the proof of Theorem 3.1, we are actually able to provide a more accurate approximation for the asymptotic parameters of the limiting joint distribution of  $(\lambda_k, \text{Tr}(\mathbf{S}_n))^\top$ . Some second-order terms which are ignored in the proof of Theorem 3.1 are sorted out in the following theorem. In applications, it may happen that the spiked eigenvalues  $\alpha'_k$ s are quite large while the sample size  $n$  remains limited. In such situations, the correction terms below will be significant although their large  $n$  limits are theoretically null. These terms are very useful for finite sample approximations as studied in the application of Section 4.

THEOREM 3.2. *The asymptotic parameters in Theorem 3.1 can be further approximated as follows:*

$$\mathbb{E}\left(\sqrt{n}\left(\frac{\lambda_k}{\psi_k} - 1\right)\right) = \frac{\alpha_k^2 \psi'(\alpha_k)}{\sqrt{n} \psi_k} \mu_{M_k} + o\left(\frac{1}{\sqrt{n} \alpha_k^2}\right),$$

$$\text{Var}\left(\sqrt{n}\left(\frac{\lambda_k}{\psi_k} - 1\right)\right) = \sum_{i=1}^m u_{ik}^4 \sigma_{\alpha_k}^2 + \sum_{i \neq j}^m u_{ik}^2 u_{jk}^2 s_{\alpha_k}^2 + \frac{\alpha_k^4}{n \psi_k^2} (\psi'(\alpha_k))^2 \sigma_{M_k}^2 + o\left(\frac{1}{n \alpha_k^2}\right),$$

$$\begin{aligned} &\text{Var}(\text{Tr}(\mathbf{S}_n) - \text{Tr}(\boldsymbol{\Sigma}_p)) \\ &= \frac{2}{n} \text{Tr}(\mathbf{V}_{p'}^2) + \frac{(\nu_4 - 3)}{n} \sum_{i=1}^{p'} [\mathbf{V}_{p'}]_{ii}^2 + \frac{1}{n} \left( \sum_{i=1}^m \Lambda_{ii}^2 (\nu_4 - 1) + \sum_{i \neq j} \Lambda_{ij}^2 \right), \end{aligned}$$

$$\begin{aligned} &\text{Cov}\left(\sqrt{n}\left(\frac{\lambda_k}{\psi_k} - 1\right), \text{Tr}(\mathbf{S}_n) - \text{Tr}(\boldsymbol{\Sigma}_p)\right) \\ &= \rho_k + \frac{y(\nu_4 - 1)}{\sqrt{n} \psi_k} \int \frac{t^2}{(1 - t/\alpha_k)^2} dH(t) + o\left(\frac{1}{\sqrt{n} \alpha_k}\right), \end{aligned}$$

where

$$\begin{aligned} \rho_k &= \frac{\alpha_k \psi'(\alpha_k)}{\sqrt{n} \psi_k} \left( (\nu_4 - 1) \sum_{i=1}^m \Lambda_{ii} u_{ik}^2 + \sum_{i \neq j} \Lambda_{ij} u_{ik} u_{jk} \right), \\ \mu_{M_k} &= -\frac{1}{\alpha_k^3} \left( \frac{y \int \frac{t^2}{(1-t/\alpha_k)^3} dH(t)}{\left(1 - \frac{y}{\alpha_k} \int \frac{t^2}{(1-t/\alpha_k)^2} dH(t)\right)^2} + \frac{y \beta_y \int \frac{t^2}{(1-t/\alpha_k)^3} dH(t)}{1 - \frac{y}{\alpha_k} \int \frac{t^2}{(1-t/\alpha_k)^2} dH(t)} \right), \\ \sigma_{M_k}^2 &= \frac{2 \underline{s}'(\psi_k) \underline{s}'''(\psi_k) - 3(\underline{s}''(\psi_k))^2}{6(\underline{s}'(\psi_k))^2} + y \beta_y (\underline{s}'(\psi_k))^2 \int \frac{t^2}{(1 + t \underline{s}(\psi_k))^4} dH(t). \end{aligned}$$

REMARK 3.1. In a related study, [28] proved, when the spiked part  $\alpha_j = \alpha_j(p) \rightarrow \infty$  while  $c_j = p/(n\alpha_j)$  is bounded and the nonspiked part

$$\frac{1}{p - m} \sum_{j=1}^{p'} \beta_{p',j} = \bar{c} + o(n^{-1/2}),$$

as  $p/n \rightarrow \infty$ , we have

$$\sqrt{n} \left\{ \frac{\lambda_j}{\alpha_j} - (1 + \bar{c} c_j + O_{\mathbb{P}}(\alpha_j^{-1} \sqrt{p/n})) \right\} \xrightarrow{d} \mathcal{N}(0, \nu_4 - 1).$$

Notice that  $\alpha_j^{-1} \sqrt{p/n} = c_j \sqrt{n/p}$ , while  $c_j$  is bounded and  $p/n \rightarrow \infty$ . Accordingly,  $O_{\mathbb{P}}(\alpha_j^{-1} \sqrt{p/n})$  is of order  $o_{\mathbb{P}}(1)$ . On the other hand, although our result in Theorem 3.1 is derived for bounded  $\alpha_j$  for ease of presentation, it turns out that it is still valid when  $\alpha_j \rightarrow \infty$ . In fact, by some simple manipulations, Theorem 3.1 implies (under our settings) that

$$\sqrt{n} \left\{ \frac{\lambda_j}{\alpha_j} - \left(1 + y \int \frac{t}{\alpha_j - t} dH(t)\right) \right\} \xrightarrow{d} \mathcal{N}(0, \nu_4 - 1).$$

It is interestingly compatible with the result in [28] where the term  $\bar{c} c_j + O_{\mathbb{P}}(\alpha_j^{-1} \sqrt{p/n})$  is equivalent to  $y \int \frac{t}{\alpha_j - t} dH(t)$  as  $\alpha_j \rightarrow \infty$ .

As an immediate application of Theorems 3.1 and 3.2, the following theorem characterizes the asymptotic behavior of an important statistic, the ratio statistic  $\frac{\lambda_k}{p^{-1} \text{Tr}(\mathbf{S}_n)}$ , which has been widely used in the literature of signal detection.

**THEOREM 3.3.** *Under Assumptions (i) to (v), for  $1 \leq k \leq m$ , we have*

$$(3.3) \quad \sqrt{n} \left( \frac{\lambda_k}{\frac{1}{p} \text{Tr}(\mathbf{S}_n)} - \frac{\psi_k}{\frac{1}{p} \text{Tr}(\mathbf{\Sigma}_p)} \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\psi_k^2}{\gamma_1^2} \left( \sum_{i=1}^m u_{ik}^2 \sigma_{\alpha_k}^2 + \sum_{i \neq j}^m u_{ik}^2 u_{jk}^2 s_{\alpha_k}^2 \right) \right),$$

where  $u_{ij}$  is the  $(i, j)$ th entry of  $\mathbf{\Lambda}$ 's eigenmatrix  $\mathbf{U}$ ,  $\gamma_1 = \int t dH(t)$ , the mean value of LSD of  $\mathbf{V}_{p'}$ ,  $\sigma_{\alpha_k}^2$  and  $s_{\alpha_k}^2$  are defined in (3.1) and (3.2).

Moreover, the asymptotic variance in (3.3) can be further approximated by

$$(3.4) \quad \begin{aligned} & \frac{\psi_k^2}{\left(\frac{1}{p} \text{Tr}(\mathbf{\Sigma}_p)\right)^2} \left( \sum_{i=1}^m u_{ik}^2 \sigma_{\alpha_k}^2 + \sum_{i \neq j}^m u_{ik}^2 u_{jk}^2 s_{\alpha_k}^2 \right) + \frac{\alpha_k^4 (\psi'(\alpha_k))^2 \sigma_{M_k}^2}{n \left(\frac{1}{p} \text{Tr}(\mathbf{\Sigma}_p)\right)^2} \\ & - \frac{2(\psi_k + \frac{\alpha_k^2 \psi'(\alpha_k)}{n} \mu_{M_k})}{p \left(\frac{1}{p} \text{Tr}(\mathbf{\Sigma}_p)\right)^3} \left[ \alpha_k \psi'(\alpha_k) \left( (v_4 - 1) \sum_{i=1}^m \Lambda_{ii} u_{ik}^2 + \sum_{i \neq j}^m \Lambda_{ij} u_{ik} u_{jk} \right) \right. \\ & \left. + y(v_4 - 1) \int \frac{t^2}{(1 - t/\alpha_k)^2} dH(t) \right] + \frac{(2\gamma_2 + (v_4 - 3)\gamma_{d,2})(\psi_k + \frac{\alpha_k^2 \psi'(\alpha_k)}{n} \mu_{M_k})^2}{p \left(\frac{1}{p} \text{Tr}(\mathbf{\Sigma}_p)\right)^4} \\ & + \frac{(\sum_{i=1}^m \Lambda_{ii}^2 (v_4 - 1) + \sum_{i \neq j} \Lambda_{ij}^2)(\psi_k + \frac{\alpha_k^2 \psi'(\alpha_k)}{n} \mu_{M_k})^2}{p^2 \left(\frac{1}{p} \text{Tr}(\mathbf{\Sigma}_p)\right)^4}. \end{aligned}$$

Under the generalized spiked population model in (1.3), by applying Delta's method to the ratio function  $f(x, y) = x/y$ , with the joint limiting distribution of  $(\lambda_1, \dots, \lambda_m, \text{Tr}(\mathbf{S}_n))^T$  in Theorems 3.1 and 3.2, we immediately have the above limiting distribution of  $\lambda_k / \text{Tr}(\mathbf{S}_n)$  (proof thus omitted). Observing that  $p^{-1} \text{Tr}(\mathbf{S}_n) - p^{-1} \text{Tr}(\mathbf{\Sigma}_p) = o_p(1)$ , one might obtain (3.3) by directly using Slutsky's theorem. However, the second-order terms in (3.4) need implementation of the Delta's method.

**4. Application.** Determination of the number of signals embedded in noise is a fundamental problem in signal processing and chemometrics community. A significant portion of this literature has been focused on the spiked covariance model arising from the following factor structure. Consider a sequence of  $p$ -dimensional random vectors  $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ , admitting a version of static  $m$ -factor structure with  $m$  fixed,

$$(4.1) \quad \mathbf{X}_t = \mathbf{A}\mathbf{F}_t + \mathbf{E}_t.$$

Here, the factors  $\mathbf{F}_t \sim \mathcal{N}(0, \mathbf{I}_m)$  are assumed to be independent of the idiosyncratic error terms  $\mathbf{E}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_p)$  with  $\sigma^2$  fixed. The loading matrix  $\mathbf{A}_{p \times m}$  is deterministic and of full rank such that  $\mathbf{A}^T \mathbf{A}$  has fixed eigenvalues  $a_1 > \dots > a_m > 0$ . Suppose that we observe  $\{\mathbf{X}_t, t = 1, \dots, n\}$  with  $n$  comparable to  $p$ . In the high dimensional context, one major inference problem in (4.1) is to infer the total number of factors  $m$ .

Note that the eigenvalues of population covariance matrix  $\mathbf{\Sigma}_p$  of  $\mathbf{X}_t$  are

$$\text{Spec}(\mathbf{\Sigma}_p) = \{a_1 + \sigma^2, \dots, a_m + \sigma^2, \underbrace{\sigma^2, \dots, \sigma^2}_{p-m}\}.$$

Thus, it is immediate to observe that  $\mathbf{X}_t$  follows the generalized spiked model (1.3) with  $\alpha_k = a_k + \sigma^2, 1 \leq k \leq m, \beta_{p',1} = \dots = \beta_{p',p'} = \sigma^2$ .



A typical approach to test the number of factors is to find all the “outliers” among eigenvalues  $\lambda_1 \geq \dots \geq \lambda_p$  of sample covariance matrix. According to the phase transition phenomenon of sample eigenvalues established in [5], it is known that the asymptotic behavior of  $\lambda_k (1 \leq k \leq m)$  differs depending on whether  $\alpha_k/\sigma^2 > 1 + \sqrt{p/n}$  or not. Only when the signal-to-noise ratio (SNR)  $\alpha_k/\sigma^2$  of the spikes is large enough can the corresponding sample eigenvalues be separated away from those bulk ones (outliers). Otherwise, these factors would be too weak and mix up with the noise. In this section, we develop a new test for presence of moderately strong factors. For any given integer  $m_0 \geq 1$  and constant  $c > 1 + \sqrt{p/n}$ , we aim to test

$$(4.2) \quad H_0 : \frac{\alpha_{m_0}}{\sigma^2} \geq c \quad \text{v.s.} \quad H_1 : \frac{\alpha_{m_0}}{\sigma^2} < c.$$

In other words, under  $H_0$ , there are at least  $m_0$  signals with SNR larger than  $c$ . If  $c = 1 + \sqrt{p/n}$ , we are actually testing for the number of moderately strong factors above the critical transition value  $1 + \sqrt{p/n}$ , that is to test

$$(4.3) \quad H_0 : m \geq m_0 \quad \text{v.s.} \quad H_1 : m < m_0.$$

Compared to [10] and [22] who have developed procedures to test “ $m = 0$ ” against “ $m > 0$ ” which still have nonzero power even when the factors are weak (SNR below the threshold  $1 + \sqrt{p/n}$ ), our test focuses more on testing existence of multiple strong spikes and their signal strength.

Since the noise variance  $\sigma^2$  is typically unknown in real applications, we propose the following normalized sample eigenvalue as our test statistic:

$$T_{m_0} = \lambda_{m_0} / \frac{1}{p - m_0} \sum_{j>m_0}^p \lambda_j$$

Here, we use  $(p - m_0)^{-1} \sum_{j>m_0}^p \lambda_j$  instead of the popular alternative  $p^{-1} \text{Tr}(\mathbf{S}_n)$  as the surrogate for  $\sigma^2$ . Although asymptotically equivalent under our conditions,  $(p - m_0)^{-1} \sum_{j>m_0}^p \lambda_j$  is often found to have better finite-sample behavior and serve as a better estimate of  $\sigma^2$  than  $p^{-1} \text{Tr}(\mathbf{S}_n)$  under cases where several large spikes or a sizable collection of medium sized spikes are present. Correspondingly,  $T_{m_0}$  often has superior detection power compared to  $\tilde{T}_{m_0} = \lambda_{m_0} / (p^{-1} \text{Tr}(\mathbf{S}_n))$  in finite sample cases.

In the literature, [21] and [13] adopted the test statistic  $\tilde{T}_1 = \lambda_1 / (p^{-1} \text{Tr}(\mathbf{S}_n))$  while they focus on testing the existence of one single spike, that is,

$$(4.4) \quad H_0 : \Sigma_p = \sigma^2 \mathbf{I}_p \quad \text{v.s.} \quad H_1 : m \geq 1.$$

Notice that this test is equivalent to the classical “sphericity test with a spiked alternative.” In the seminal paper [16], it has been proven that when  $\Sigma_p = \sigma^2 \mathbf{I}_p$  with Gaussian data, as  $p, n \rightarrow \infty$  with  $p/n \rightarrow y$ ,

$$\frac{\lambda_1/\sigma^2 - \mu_{np}}{\sigma_{np}} \xrightarrow{d} \text{TW}_1,$$

where  $\mu_{np} = \frac{1}{n}(\sqrt{n-1} + \sqrt{p})^2$ ,  $\sigma_{np} = \sqrt{\frac{\mu_{np}}{n}}(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}})^{1/3}$  and  $\text{TW}_1$  denotes Tracy–Widom distribution of order 1. Ma [20] further refined the centering and scaling parameters which improves the convergence rate from  $O((n \wedge p)^{-1/3})$  to  $O((n \wedge p)^{-2/3})$ . Note that the fluctuation in  $\hat{\sigma}^2 = p^{-1} \text{Tr}(\mathbf{S}_n)$  is of order  $O(1/p)$ , which is negligible compared to that in  $\lambda_1$ . Therefore, we still have, as  $p, n \rightarrow \infty$ ,

$$\mathbb{P}\left(\frac{\tilde{T}_1 - \mu_{np}}{\sigma_{np}} < s\right) \rightarrow \text{TW}_1(s).$$

But in finite samples, the quality of this approximation for  $\tilde{T}_1$  breaks down due to studentization. Nadler [21] derived an explicit approximation formula for the tail probabilities of  $\tilde{T}_1$ , which provides a significantly better fit to the empirical distribution of  $\tilde{T}_1$ . But still, it is not a proper distribution function. Deo [13] suggested an alternative variance adjustment for the scaling parameter  $\sigma_{np}$  to improve the finite sample performance of  $\tilde{T}_1$ . Tests based on these two corrections eliminate the downward size bias of the uncorrected test and improve its power performance for small values of  $(p, n)$ .

However, the problem of testing the existence of multiple spikes has not been fully resolved yet, that is, for  $m_0 > 1$ , to test

$$(4.5) \quad H_0 : m \leq m_0 \quad \text{v.s.} \quad H_1 : m > m_0.$$

First, in the presence of  $m$  ( $m > 1$ ) spikes, the conjecture of

$$(4.6) \quad \frac{\lambda_{m+1}/\sigma^2 - \mu_{n,p-m}}{\sigma_{n,p-m}} \xrightarrow{d} \text{TW}_1$$

has not been proven yet except for the complex Gaussian case (GUE case) in [5]. Second, the finite sample bias caused by the replacement of  $\sigma^2$  with  $p^{-1} \text{Tr}(\mathbf{S}_n)$  becomes more severe under  $H_0$  in (4.5). Deo [13] and Kritchman and Nadler [18] adopted alternative estimators of  $\sigma^2$  based on  $p - m$  bulk eigenvalues. However, their testing procedures are still based on the unverified conjecture (4.6) and simulation experiments show that these tests are still uniformly undersized due to the negative bias in the presence of spikes. Further analytic tools are needed to correct this bias.

In this paper, we start from a different perspective by studying the behavior of  $\lambda_m / (\frac{1}{p-m} \sum_{j>m}^p \lambda_j)$  instead of  $\lambda_{m+1}$  in the presence of  $m$  spikes and aim to test (4.2) and (4.3). Although our test hypothesis is different from previous works, it can still be used to determine the total number of factors by performing a sequence of hypothesis tests on testing whether  $\lambda_k$  ( $1 \leq k \leq m$ ) arise from the signal or noise. The limiting distribution of our test statistic  $T_m$  is fully implementable under  $H_0$  in (4.2) assuming that all the conditions in Theorems 3.1 and 3.2 are satisfied. Our test statistic is not only capable of testing the existence of multiple spikes, but can also be used to test their signal strength. Higher order corrections are further made to alleviate finite sample bias, which ensures satisfactory testing size and power even when  $(p, n)$  is not large.

The corollary below follows from a direct implementation of Theorem 3.2.

**COROLLARY 4.1.** *Considering the factor model (4.1), for  $1 \leq k \leq m$ , as  $n \rightarrow \infty$ ,  $p = p_n \rightarrow \infty$  such that  $p/n \rightarrow y > 0$ , we have*

$$(4.7) \quad \sqrt{n} \left( \frac{\lambda_k}{\frac{1}{p-k} \sum_{j>k} \lambda_j} - \frac{\psi_k}{\tilde{\sigma}^2} \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{2\alpha_k^2 \psi'(\alpha_k)}{\sigma^4} \right),$$

where

$$\psi_j = \alpha_j + \frac{y\alpha_j\sigma^2}{\alpha_j - \sigma^2}, \quad \psi'(\alpha_j) = 1 - \frac{y\sigma^4}{(\alpha_j - \sigma^2)^2}, \quad \alpha_j = a_j + \sigma^2,$$

$$\tilde{\sigma}^2 = \frac{1}{p-k} \left[ \text{Tr}(\Sigma_p) - \sum_{j=1}^k \left\{ \psi_j + \frac{\alpha_j^2 \psi'(\alpha_j)}{n} \mu_{M_j} \right\} \right].$$

Additionally, we have a refined correction for the variance of  $\frac{\lambda_k}{\frac{1}{p-k} \sum_{j>k} \lambda_j}$ ,

$$\begin{aligned}
 \sigma_{*,k}^2 &= \frac{2\alpha_k^2 \psi'(\alpha_k)}{\tilde{\sigma}^4} + \frac{4\alpha_k^2 \psi'(\alpha_k) \{ \psi_k + \frac{2\alpha_k^2 \psi'_k(\alpha_k)}{n} \mu_{M_k} \}}{(p-k)\tilde{\sigma}^6} \\
 &+ \frac{\alpha_k^4 (\psi'(\alpha_k))^2 \sigma_{M_k}^2}{n\tilde{\sigma}^4} - \frac{4\alpha_k^2 \{ \psi_k + \frac{\alpha_k^2 \psi'_k(\alpha_k)}{n} \mu_{M_k} \}}{(p-k)\tilde{\sigma}^6} \\
 &+ \frac{n}{(p-k)^2} \left( 2y\sigma^4 + \frac{2}{n} \sum_{i=1}^m \alpha_i^2 \right) \frac{\{ \psi_k + \frac{2\alpha_k^2 \psi'_k(\alpha_k)}{n} \mu_{M_k} \}^2}{\tilde{\sigma}^8} \\
 &+ \frac{2\psi_k \alpha_k^4 (\psi'(\alpha_k))^2 \sigma_{M_k}^2}{n(p-k)\tilde{\sigma}^6} + \frac{\psi_k^2 \sum_{j=1}^k \{ 2\alpha_j^2 \psi'(\alpha_j) - 4\alpha_j^2 \}}{(p-k)^2 \tilde{\sigma}^8}.
 \end{aligned}
 \tag{4.8}$$

Here,

$$\mu_{M_j} = -\frac{y\sigma^4}{(\alpha_j - \sigma^2)^3 \left\{ 1 - \frac{y\sigma^4}{(\alpha_j - \sigma^2)^2} \right\}^2}, \quad \sigma_{M_j}^2 = \frac{2\underline{s}'(\psi_j) \underline{s}'''(\psi_j) - 3(\underline{s}''(\psi_j))^2}{6(\underline{s}'(\psi_j))^2}.$$

Let  $Z_\alpha$  be the lower- $\alpha$  quantile of the standard normal distribution at level  $\alpha$ . In order to define an appropriate critical value, one notes that the null hypothesis (4.2) is composite. For a given value of  $\alpha_{m_0}$  under the null, the decision rule is to

$$\text{reject } H_0 \text{ in (4.2) at the value } \alpha_{m_0} \text{ if } T_{m_0} < \frac{\psi_{m_0}}{\tilde{\sigma}^2} + Z_\alpha \cdot \frac{\sigma_{*,m_0}}{\sqrt{n}}.$$

Letting  $t_k = \alpha_k/\sigma^2$ , both the critical value above and the refined variance (4.8) can be expressed as functions of these  $t_k$ 's:

$$\begin{aligned}
 q_{n,\alpha} &= q_{n,\alpha}(t_{m_0}, t_{m_0-1}, \dots, t_1) = \frac{\psi_{m_0}}{\tilde{\sigma}^2} + Z_\alpha \cdot \frac{\sigma_{*,m_0}}{\sqrt{n}} \\
 &= \frac{t_{m_0} + \frac{y}{1-1/t_{m_0}}}{1 - \frac{1}{p-m_0} \sum_{j=1}^{m_0} \frac{y}{1-1/t_j}} + Z_\alpha \cdot \frac{\sigma_{*,m_0}}{\sqrt{n}}, \\
 \sigma_{*,m_0}^2 &= \sigma_{*,m_0}^2(t_{m_0}, t_{m_0-1}, \dots, t_1) \\
 &= 2t_{m_0}^2 \left( 1 - \frac{y}{(t_{m_0} - 1)^2} \right) \left( \frac{1}{1 - \frac{1}{p-m_0} \sum_{j=1}^{m_0} \frac{y}{1-1/t_j}} \right)^2 \\
 &\quad - \frac{4yt_{m_0}^2}{(p-m_0)(t_{m_0} - 1)^2} \left( t_{m_0} + \frac{y}{1-1/t_{m_0}} \right) \left( \frac{1}{1 - \frac{1}{p-m_0} \sum_{j=1}^{m_0} \frac{y}{1-1/t_j}} \right)^3 \\
 &\quad + \frac{2yn}{(p-m_0)^2} \left( t_{m_0} + \frac{y}{1-1/t_{m_0}} \right)^2 \left( \frac{1}{1 - \frac{1}{p-m_0} \sum_{j=1}^{m_0} \frac{y}{1-1/t_j}} \right)^4 \\
 &\quad + \frac{t_{m_0}^2 (1 - \frac{y}{(t_{m_0} - 1)^2})^2}{n} \left( \frac{1}{1 - \frac{1}{p-m_0} \sum_{j=1}^{m_0} \frac{y}{1-1/t_j}} \right)^2 \left\{ \frac{4yt_{m_0}}{3(t_{m_0} - 1)^3} \right.
 \end{aligned}$$

$$\begin{aligned}
 & - \frac{4yt_{m_0}}{3(t_{m_0} - 1)^3 \left(1 - \frac{y}{(t_{m_0}-1)^2}\right)^3} + \frac{2y^2t_{m_0}^2}{3(t_{m_0} - 1)^6 \left(1 - \frac{y}{(t_{m_0}-1)^2}\right)^4} \\
 & + \left. \frac{2yt_{m_0}^2}{(t_{m_0} - 1)^4} + \frac{4y^2t_{m_0}^2}{3(t_{m_0} - 1)^6 \left(1 - \frac{y}{(t_{m_0}-1)^2}\right)^2} \right\}.
 \end{aligned}$$

Therefore, for the composite null (4.2), we will use the critical value

$$(4.9) \quad q_{n,\alpha}^* = q_{n,\alpha}^*(t_{m_0-1}, \dots, t_1) = \inf_{c \leq t_{m_0} < t_{m_0-1}} q_{n,\alpha}(t_{m_0}, t_{m_0-1}, \dots, t_1),$$

and reject the null if  $T_{m_0} < q_{n,\alpha}^*$ . For this procedure, it holds that

$$\limsup_{n \rightarrow \infty} \sup_{c \leq t_{m_0} < t_{m_0-1}} \mathbb{P}(T_{m_0} < q_{n,\alpha}^*) \leq \alpha.$$

Implementation of this procedure finally requires to estimate the values of larger spikes  $\{t_k, 1 \leq k < m_0\}$  that appear in the critical value functions  $q_{n,\alpha}$  and  $q_{n,\alpha}^*$ . As a matter of fact, consistent estimates for these spike values have been proposed in [1]. It is known that, under  $H_0$  in (4.2), for distant spikes  $\alpha_k$  ( $1 \leq k \leq m_0$ ), as  $p, n \rightarrow \infty, p/n \rightarrow y, \lambda_k \xrightarrow{\text{a.s.}} \ell_k$ , satisfying  $\underline{s}(\ell_k) = -\frac{1}{\alpha_k}$ . Based on this result, [1] proposed a consistent estimator  $\hat{\alpha}_k$  for  $\alpha_k$ ,

$$\hat{\alpha}_k = -\frac{1}{\hat{\underline{s}}(\ell_k)}, \quad \hat{\underline{s}}(\ell_k) = -\frac{1-y}{\lambda_k} + \frac{1}{n} \sum_{i>m_0} \frac{1}{\lambda_i - \lambda_k}.$$

Here, the noise level  $\sigma^2$  is intrinsically hidden inside the values of sample eigenvalues  $\lambda_j$ 's, not explicitly used in the estimation of  $\alpha_k$ . Note that the conditions used in [1] which ensure the consistency of  $\hat{\alpha}_k$  are satisfied under our settings (Assumptions (i) to (v)). Thus for  $1 \leq k < m_0$ , the  $t_k$ 's can be consistently estimated by

$$\hat{t}_k = \frac{\hat{\alpha}_k}{(\text{Tr}(\mathbf{S}_n) - \sum_{\ell=1}^{m_0} \hat{\alpha}_\ell) / (p - m_0)}.$$

Plugging these estimates into the critical value functions  $q_{n,\alpha}$  and  $q_{n,\alpha}^*$  leads to a full implementation of the test.

To summarize, the proposed testing procedure is to

$$(4.10) \quad \text{reject } H_0 \text{ in (4.2) if } T_{m_0} < \hat{q}_{n,\alpha}^*,$$

where

$$(4.11) \quad \hat{q}_{n,\alpha}^* = q_{n,\alpha}^*(\hat{t}_{m_0-1}, \dots, \hat{t}_1) = \inf_{c \leq t_{m_0} < \hat{t}_{m_0-1}} q_{n,\alpha}(t_{m_0}, \hat{t}_{m_0-1}, \dots, \hat{t}_1).$$

Simulation experiments are conducted in the Supplementary Material [19] to examine the performance of out testing procedure. Empirical data are generated following the factor model in (4.1). Different numbers of factors and signal strength are assigned for various model settings. Results (Tables 1 and 2) show that our test works for different numbers of factors. It has better performance for higher signal strength levels. The same supplement section has also designed experiments to illustrate the necessity of incorporating the second-order correction to the asymptotic variance proposed in (4.8). Numerical comparison is made between (4.10) and the following testing procedure to

$$(4.12) \quad \text{reject } H_0 \text{ in (4.2) if } T_{m_0} < \tilde{q}_{n,\alpha}^*,$$

where  $\tilde{q}_{n,\alpha}^*$  is defined as  $\hat{q}_{n,\alpha}^*$  in (4.11) except that  $\frac{\psi_{m_0}}{\sigma^2}$  is replaced by  $\frac{\psi_{m_0}}{\sigma^2}$  and the refined asymptotic variance  $\sigma_{*,m_0}^2$  used there is replaced by  $\tilde{\sigma}_{m_0}^2 = 2t_{m_0}^2 - \frac{2yt_{m_0}^2}{(t_{m_0}-1)^2}$ . Tables 3 and 4 in

the Supplementary Material [19] show that this refined correction (4.8) for the variance plays an important role in controlling the size and improves the power in the testing procedure, especially when the data dimension and sample size are relatively small.

**5. Proofs.**

5.1. *Proof of Theorem 3.1.* Considering the block structure of population covariance matrix  $\Sigma_p$ , the analysis is carried out using a decomposition into blocks of size  $m$  and  $p - m$ , respectively. Define

$$Y_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix}, \quad \Sigma_p = \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{p'} \end{pmatrix}, \quad X_i = \begin{pmatrix} X_{1i} \\ X_{2i} \end{pmatrix} = \begin{pmatrix} \Lambda^{1/2} Y_{1i} \\ \mathbf{V}_{p'}^{1/2} Y_{2i} \end{pmatrix}.$$

The sample covariance matrix is then

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^T = \frac{1}{n} \begin{pmatrix} X_1 X_1^T & X_1 X_2^T \\ X_2 X_1^T & X_2 X_2^T \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

where

$$\begin{aligned} S_{11} &= \frac{1}{n} \Lambda^{1/2} Y_1 Y_1^T \Lambda^{1/2}, & S_{22} &= \frac{1}{n} \mathbf{V}_{p'}^{1/2} Y_2 Y_2^T \mathbf{V}_{p'}^{1/2}, \\ X_1 &= (X_{11}, \dots, X_{1n}), & X_2 &= (X_{21}, \dots, X_{2n}), \\ Y_1 &= (Y_{11}, \dots, Y_{1n}), & Y_2 &= (Y_{21}, \dots, Y_{2n}). \end{aligned}$$

The proof of Theorem 3.1 builds on the above block-decomposition analysis of spiked and nonspiked ones.  $\text{Tr}(S_n)$  follows the decomposition  $\text{Tr}(S_n) = \text{Tr}(S_{11}) + \text{Tr}(S_{22})$ . It will later be shown that  $\text{Tr}(S_{22})$  is asymptotically independent of the random vector  $(\lambda_1, \dots, \lambda_m)^T$ , while the covariance between  $\text{Tr}(S_{11})$  and  $(\lambda_1, \dots, \lambda_m)^T$  is of the order  $O(1/\sqrt{n})$ . The proof in general consists of four steps as follows:

- Step 1. deriving the asymptotic joint distribution of  $(\lambda_1, \dots, \lambda_m)^T$ ;
- Step 2. deriving the marginal limiting distribution of  $\text{Tr}(S_{11})$  and  $\text{Tr}(S_{22})$ ;
- Step 3. deriving the asymptotic joint distribution of  $((\lambda_k)_{1 \leq k \leq m}, \text{Tr}(S_n))^T$ .

*Step 1: Joint limiting distribution of  $(\lambda_1, \dots, \lambda_m)^T$ .*

Many efforts in the literature have been put into the study of the asymptotic behavior of extreme sample eigenvalues under various spiked population models. Notably, [3] derived a CLT for sample eigenvalues corresponding to one distant spiked eigenvalue under a more generalized model where possible multiplicity of spiked eigenvalues is allowed such that

$$(5.1) \quad \text{Spec}(\Sigma_p) = \{\underbrace{\alpha_1, \dots, \alpha_1}_{m_1}, \dots, \underbrace{\alpha_K, \dots, \alpha_K}_{m_K}, \beta_{p',1}, \dots, \beta_{p',p'}\}.$$

Here, we eliminate the multiplicity of spikes in model (5.1) and focus on the correlation among sample eigenvalues corresponding to different spikes. Wang et al. [26] studied such correlations under the model where

$$\text{Spec}(\Sigma_p) = \{\underbrace{\alpha_1, \dots, \alpha_1}_{m_1}, \dots, \underbrace{\alpha_K, \dots, \alpha_K}_{m_K}, 1, \dots, 1\},$$

while our model (1.3) is more general in the sense of bulk eigenvalue distribution  $H(t)$ . The proof in general combines the Z-estimation scheme in [3] and the result of joint CLT for several random sesquilinear forms in [26]. Detailed proofs are presented here, which will also be used in subsequent steps and the proof of Theorem 3.2.

Noting that, whenever  $\mathbf{A}$  is invertible,

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}| \cdot |\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}|,$$

an eigenvalue  $\lambda_i$  of  $\mathbf{S}_n$  that is not an eigenvalue of  $\mathbf{S}_{22}$  satisfies

$$(5.2) \quad 0 = |\lambda_i \mathbf{I}_p - \mathbf{S}_n| = |\lambda_i \mathbf{I}_{p-m} - \mathbf{S}_{22}| \cdot |\lambda_i \mathbf{I}_m - \mathbf{K}_n(\lambda_i)|,$$

where

$$\mathbf{K}_n(\ell) := \mathbf{S}_{11} + \mathbf{S}_{12}(\ell \mathbf{I}_{p-m} - \mathbf{S}_{22})^{-1} \mathbf{S}_{21}.$$

Thus, the eigenvalues of  $\mathbf{S}_n$  satisfy

$$(5.3) \quad |\lambda_i \mathbf{I}_m - \mathbf{K}_n(\lambda_i)| = 0.$$

Consider a real number  $\ell$  outside the support of LSD  $F^{y,H}$  of  $\mathbf{S}_{22}$  and the goal is to find the limit of random matrix  $\mathbf{K}_n(\ell)$  with fixed dimension  $m$ . Since for  $\ell \neq 0$  which is not an eigenvalue of  $\mathbf{A}^\top \mathbf{A}$ ,

$$\mathbf{I}_n + \mathbf{A}(\ell \mathbf{I}_{p-m} - \mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top = \ell(\ell \mathbf{I}_n - \mathbf{A}\mathbf{A}^\top)^{-1},$$

it holds that

$$(5.4) \quad \begin{aligned} \mathbf{K}_n(\ell) &= \frac{1}{n} \mathbf{X}_1 \left\{ \mathbf{I}_n + \frac{1}{n} \mathbf{X}_2^\top \left( \ell \mathbf{I}_{p-m} - \frac{1}{n} \mathbf{X}_2 \mathbf{X}_2^\top \right)^{-1} \mathbf{X}_2 \right\} \mathbf{X}_1^\top \\ &= \frac{\ell}{n} \mathbf{\Lambda}^{1/2} \mathbf{Y}_1 \left( \ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \mathbf{Y}_1^\top \mathbf{\Lambda}^{1/2}. \end{aligned}$$

By Assumptions (i)–(iv),  $\ell$  is outside the support of LSD  $F^{y,H}$  of  $\mathbf{S}_{22}$  and for  $n$  large enough, the operator norm of  $(\ell \mathbf{I}_{p-m} - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2)^{-1}$  is bounded. Meanwhile,  $\mathbf{Y}_1$  is independent of  $\mathbf{Y}_2$ . Then by the law of large numbers, Corollary 3.9 in [17], and Theorem 11.8 in [29], we have, as  $p, n \rightarrow \infty, p/n \rightarrow y > 0$ ,

$$\mathbf{\Lambda}^{-1/2} \mathbf{K}_n(\ell) \mathbf{\Lambda}^{-1/2} \xrightarrow{\text{a.s.}} -\ell \underline{s}(\ell) \cdot \mathbf{I}_m,$$

where  $\underline{s}(\ell)$  is the Stieltjes transform of the LSD of  $\frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2$ .

Therefore, the eigenvalue  $\lambda_i$  of  $\mathbf{S}_n$  satisfying (5.3) converges almost surely to the limit  $\ell$  such that

$$|\ell \cdot \mathbf{\Lambda}^{-1} + \ell \underline{s}(\ell) \cdot \mathbf{I}_m| = 0,$$

that is,

$$\underline{s}(\ell) = -1/\alpha_k, \quad \ell = \psi(\alpha_k), k = 1, \dots, m,$$

where  $\text{Spec}(\mathbf{\Lambda}) = \{\alpha_1, \dots, \alpha_m\}$ . The following lemma, due to [25], characterizes the close relationship between the supports of the generating measure  $H$  and the Marčenko–Pastur (M-P) distribution  $F^{y,H}$ .

LEMMA 5.1 (By [25]). *If  $\lambda \notin \Gamma_{F^{y,H}}$ , then  $\underline{s}(\lambda) \neq 0$  and  $\alpha = -1/\underline{s}(\lambda)$  satisfies:*

- (1)  $\alpha \notin \Gamma_H$  and  $\alpha \neq 0$  (so that  $\psi(\alpha)$  is well defined);
- (2)  $\psi'(\alpha) > 0$ .

Conversely, if  $\alpha$  satisfies (1)–(2), then  $\lambda = \psi(\alpha) \notin \Gamma_{F^{y,H}}$ .

By Lemma 5.1,  $\ell = \psi(\alpha_k)$  is outside the support of LSD  $F^{y,H}$  if and only if  $\psi'(\alpha_k) > 0$ . By Assumption (v), all spiked values  $\alpha_k$  are large enough to make  $\psi'(\alpha_k) > 0$ . Therefore, the limits

$$\ell = \psi(\alpha_k) =: \psi_k, \quad k = 1, \dots, m$$

are all outside the support of LSD  $F^{y,H}$ . Meanwhile, for the  $m$  largest eigenvalues  $\lambda_1, \dots, \lambda_m$  of  $\mathbf{S}_n$ , as  $p, n \rightarrow \infty, p/n \rightarrow y > 0$ ,

$$\frac{\lambda_k}{\psi_k} \xrightarrow{\text{a.s.}} 1 \quad \text{for } 1 \leq k \leq m.$$

Note that  $\psi(\cdot)$  is the functional inverse of function  $\alpha : x \mapsto -1/\underline{s}(x)$ , then

$$\begin{aligned} \underline{s}(\psi_k) &= -\frac{1}{\alpha_k}, & \underline{s}'(\psi_k) &= \frac{1}{\alpha_k^2 \psi'(\alpha_k)}, \\ \underline{s}''(\psi_k) &= -\frac{2}{\alpha_k^3 (\psi'(\alpha_k))^2} - \frac{\psi''(\alpha_k)}{\alpha_k^2 (\psi'(\alpha_k))^3}, \\ \underline{s}'''(\psi_k) &= \frac{6}{\alpha_k^4 (\psi'(\alpha_k))^3} - \frac{2\psi''(\alpha_k)\underline{s}''(\psi_k)}{(\psi'(\alpha_k))^2} \\ &\quad - \frac{\alpha_k^2 \psi'(\alpha_k) \psi'''(\alpha_k) - \psi''(\alpha_k)(2\alpha_k \psi'(\alpha_k) + \alpha_k^2 \psi''(\alpha_k))}{\alpha_k^4 (\psi'(\alpha_k))^2}, \end{aligned} \tag{5.5}$$

where  $[\cdot]'|_{\alpha=\alpha_k}$  means to take first-order derivative with regard to  $\alpha$  and then let  $\alpha = \alpha_k$ . Here,

$$\begin{aligned} \psi_k = \psi(\alpha_k) &= \alpha_k + y \int \frac{t}{1 - t/\alpha_k} dH(t), \\ \psi'(\alpha_k) &= 1 - y \int \frac{t^2}{(\alpha_k - t)^2} dH(t), & \psi''(\alpha_k) &= 2y \int \frac{t^2}{(\alpha_k - t)^3} dH(t). \end{aligned}$$

Denote  $\underline{s}_n(\ell) = \frac{1}{n} \text{Tr}(\ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2)^{-1}$ , then

$$\mathbf{R}_n(\ell) = \frac{\ell}{\sqrt{n}} \mathbf{\Lambda}^{1/2} \mathbf{R}_n(\ell) \mathbf{\Lambda}^{1/2} - \ell \mathbf{\Lambda} \underline{s}(\ell) + \ell \mathbf{\Lambda} (\underline{s}(\ell) - \underline{s}_n(\ell)),$$

where  $\mathbf{R}_n(\ell)$  is a sequence of  $m \times m$  random matrix-valued processes

$$\begin{aligned} \mathbf{R}_n(\ell) &= \frac{1}{\sqrt{n}} \left( \mathbf{Y}_1 \left( \ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \mathbf{Y}_1^\top \right. \\ &\quad \left. - \mathbf{I}_m \text{Tr} \left( \ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \right), \quad \ell \in \mathcal{U}. \end{aligned} \tag{5.6}$$

Here,  $\mathcal{U}$  is a compact set of indexes outside the support of LSD of  $\mathbf{S}_{22}$ .

The establishment of CLT for extreme sample eigenvalues  $\lambda'_i s$  relies heavily on the finite dimensional convergence of processes

$$\{\mathbf{R}_n(\ell), \ell \in \mathcal{U}\} \quad \text{and} \quad \{n(\underline{s}_n(z) - \underline{s}(z)), z \in \mathbb{C} \setminus \Gamma_{F^{y,H}}\},$$

which has been well established in [3], Lemma 1.1 in [4]. More specifically, we have the following lemma.

LEMMA 5.2 (By Theorem 11.10 in [29]). *Under Assumptions (i) to (iv), for any  $L$  index values  $\{\ell_j\}$ , the  $L$  random matrices*

$$\{\mathbf{R}_n(\ell_1), \dots, \mathbf{R}_n(\ell_L)\}$$

weakly converge to  $L$  Gaussian random matrices determined as follows: for arbitrary  $L$  numbers  $a_1, \dots, a_L$ , the random matrix

$$\tilde{\mathbf{R}}_n = a_1 \mathbf{R}_n(\ell_1) + \dots + a_L \mathbf{R}_n(\ell_L)$$

weakly converges to a Gaussian random matrix  $\mathbf{R} = \{R_{ij}\}$  where:

(1) the diagonal entries are i.i.d. zero-mean Gaussian with variance

$$\text{Var}(R_{ii}) = w(\mathbb{E}|Y_{ij}|^4 - 3) + 2\theta;$$

(2) the upper off-diagonal entries are i.i.d. zero-mean Gaussian with variance  $\theta$ ;

(3) all these entries are mutually independent.

Here, the parameters  $\theta$  and  $w$  are

$$\theta = \sum_{j=1}^L a_j^2 \underline{s}'(\ell_j) + 2 \sum_{j < k} a_j a_k \frac{\underline{s}(\ell_j) - \underline{s}(\ell_k)}{\ell_j - \ell_k}, \quad w = \left( \sum_{j=1}^L a_j \underline{s}(\ell_j) \right)^2.$$

Notice that

$$\lambda_k \mathbf{I}_m - \mathbf{K}_n(\lambda_k) = \lambda_k \mathbf{I}_m - \frac{\lambda_k}{\sqrt{n}} \mathbf{\Lambda}^{1/2} \mathbf{R}_n(\lambda_k) \mathbf{\Lambda}^{1/2} + \lambda_k \mathbf{\Lambda} \underline{s}(\lambda_k) - \lambda_k \mathbf{\Lambda} (\underline{s}(\lambda_k) - \underline{s}_n(\lambda_k)).$$

Since we have spectral decomposition  $\mathbf{\Lambda} = \mathbf{U} \text{diag}(\alpha_1, \dots, \alpha_m) \mathbf{U}^\top$ ,

$$\begin{aligned} & \mathbf{U}^\top (\mathbf{I}_m - \lambda_k^{-1} \mathbf{K}_n(\lambda_k)) \mathbf{U} \\ &= \begin{pmatrix} \ddots & & \\ & (1 + \alpha_u \underline{s}(\lambda_k)) & \\ & & \ddots \end{pmatrix}_{m \times m} \\ & - \frac{1}{\sqrt{n}} \mathbf{U}^\top \mathbf{\Lambda}^{1/2} \mathbf{R}_n(\lambda_k) \mathbf{\Lambda}^{1/2} \mathbf{U} - \begin{pmatrix} \ddots & & \\ & \alpha_u (\underline{s}(\lambda_k) - \underline{s}_n(\lambda_k)) & \\ & & \ddots \end{pmatrix}_{m \times m}. \end{aligned}$$

Now considering  $\delta_{n,k} = \sqrt{n}(\frac{\lambda_k}{\psi_k} - 1)$ ,  $\psi_k = \alpha_k + y \int \frac{t}{1-t/\alpha_k} dH(t)$ , by Taylor expansion, we have

$$\underline{s}(\lambda_k) = \underline{s}(\psi_k) + \frac{\psi_k}{\sqrt{n}} \cdot \delta_{n,k} \cdot \underline{s}'(\psi_k) + O_{\mathbb{P}}\left(\frac{1}{n}\right),$$

which then yields

$$\begin{aligned} & \mathbf{U}^\top (\mathbf{I}_m - \lambda_k^{-1} \mathbf{K}_n(\lambda_k)) \mathbf{U} \\ &= \begin{pmatrix} \ddots & & \\ & 1 + \alpha_u \underline{s}(\psi_k) + \alpha_u \left( \frac{\psi_k}{\sqrt{n}} \cdot \delta_{n,k} \cdot \underline{s}'(\psi_k) + O_{\mathbb{P}}\left(\frac{1}{n}\right) \right) & \\ & & \ddots \end{pmatrix} \\ & - \frac{1}{\sqrt{n}} \mathbf{U}^\top \mathbf{\Lambda}^{1/2} \mathbf{R}_n(\lambda_k) \mathbf{\Lambda}^{1/2} \mathbf{U} - \begin{pmatrix} \ddots & & \\ & \alpha_u (\underline{s}(\lambda_k) - \underline{s}_n(\lambda_k)) & \\ & & \ddots \end{pmatrix}. \end{aligned}$$



First, it can be seen that all the nondiagonal terms tend to zero on the right-hand side. Then for a diagonal term with index  $u \neq k$ , by definition  $1 + \underline{s}(\psi_k)\alpha_u \neq 0$  and it is indeed the leading term with the remaining three terms converging to zero. As for the  $k$ th diagonal term,  $1 + \underline{s}(\psi_k)\alpha_k = 0$  by definition, thus the  $k$ th diagonal term reduces to

$$\alpha_k \left( \frac{\psi_k}{\sqrt{n}} \cdot \delta_{n,k} \cdot \underline{s}'(\psi_k) + O_{\mathbb{P}}\left(\frac{1}{n}\right) \right) - \frac{1}{\sqrt{n}} [\mathbf{U}^{\top} \mathbf{\Lambda}^{1/2} \mathbf{R}_n(\lambda_k) \mathbf{\Lambda}^{1/2} \mathbf{U}]_{kk} - \alpha_k (\underline{s}(\lambda_k) - \underline{s}_n(\lambda_k)).$$

Noting that  $|\mathbf{U}^{\top} (\mathbf{I}_m - \lambda_k^{-1} \mathbf{K}_n(\lambda_k)) \mathbf{U}| = 0$ , for  $n$  sufficiently large, we have

$$(5.7) \quad \left| \frac{\psi_k}{\sqrt{n}} \cdot \delta_{n,k} \cdot \underline{s}'(\psi_k) + O_{\mathbb{P}}\left(\frac{1}{n}\right) - \frac{1}{\sqrt{n}} [\mathbf{U}^{\top} \mathbf{R}_n(\lambda_k) \mathbf{U}]_{kk} - (\underline{s}(\lambda_k) - \underline{s}_n(\lambda_k)) \right| = 0.$$

Taking into account the convergence of process  $\{\mathbf{R}_n(\ell), \ell \in \mathcal{U}\}$  and  $\{M_n(z) = n(\underline{s}_n(z) - \underline{s}(z))\}$  in Lemma 5.2 and [4], it follows that  $\delta_{n,k}$  weakly converges to a solution of the limit of

$$\left| \frac{\psi_k}{\sqrt{n}} \cdot \delta_{n,k} \cdot \underline{s}'(\psi_k) - \frac{1}{\sqrt{n}} [\mathbf{U}^{\top} \mathbf{R}_n(\psi_k) \mathbf{U}]_{kk} - O_{\mathbb{P}}\left(\frac{1}{n}\right) \right| = 0,$$

that is,

$$\delta_{n,k} = \sqrt{n} \left( \frac{\lambda_k}{\psi_k} - 1 \right) \xrightarrow{d} \frac{[\mathbf{U}^{\top} \mathbf{R}(\psi_k) \mathbf{U}]_{kk}}{\psi_k \underline{s}'(\psi_k)}.$$

Denote

$$\mathbf{G}(\psi_k) = \frac{\mathbf{R}(\psi_k)}{\psi_k \underline{s}'(\psi_k)}.$$

Then by Lemma 5.2,  $\mathbf{G}(\psi_k)$  is a Gaussian random matrix with mutually independent entries where the diagonal entries are i.i.d. zero-mean Gaussian with variance

$$\text{Var}([\mathbf{G}(\psi_k)]_{ii}) = \frac{(\nu_4 - 3)\underline{s}(\psi_k)^2 + 2\underline{s}'(\psi_k)}{(\underline{s}'(\psi_k))^2 \psi_k^2},$$

and the upper off-diagonal entries are i.i.d. zero mean Gaussian with variance

$$\text{Var}([\mathbf{G}(\psi_k)]_{ij}) = \frac{1}{\underline{s}'(\psi_k) \psi_k^2}.$$

Now we consider the asymptotic joint distribution of  $(\lambda_{k_1}, \lambda_{k_2})^{\top}$ ,  $1 \leq k_1 \neq k_2 \leq m$ . It can be seen from the previous proof that the leading term of  $\delta_{n,k} = \sqrt{n}(\frac{\lambda_k}{\psi_k} - 1)$  is  $\frac{[\mathbf{U}^{\top} \mathbf{R}_n(\psi_k) \mathbf{U}]_{kk}}{\underline{s}'(\psi_k) \psi_k}$ . Thus the correlation between limits of  $(\lambda_{k_1}, \lambda_{k_2})^{\top}$  is determined by the joint limiting distribution of the two random sesquilinear forms  $\mathbf{R}_n(\psi_{k_1})$  and  $\mathbf{R}_n(\psi_{k_2})$ . This task is nontrivial. Here, we apply a joint CLT for random vector whose components are function of random sesquilinear forms by [27].

LEMMA 5.3 (By [27]). *Consider a sequence  $(\mathbf{X}_i, \mathbf{Y}_i)_{i \in \mathbb{N}}$  of i.i.d. real valued zero-mean random vectors belonging to  $\mathbb{R}^K \times \mathbb{R}^K$  with finite fourth-order moment:*

$$\begin{aligned} \mathbf{X}_i &= (X_{1i}, \dots, X_{Ki})^{\top}, & 1 \leq i \leq n, & & \mathbf{X}(\ell) &= (X_{\ell 1}, \dots, X_{\ell n})^{\top}, & 1 \leq \ell \leq K, \\ \mathbf{Y}_i &= (Y_{1i}, \dots, Y_{Ki})^{\top}, & 1 \leq i \leq n, & & \mathbf{Y}(\ell) &= (Y_{\ell 1}, \dots, Y_{\ell n})^{\top}, & 1 \leq \ell \leq K, \end{aligned}$$

and  $\rho(\ell) := \mathbb{E}(X_{\ell 1} Y_{\ell 1})$ . Let  $\{\mathbf{A}_n\}$  and  $\{\mathbf{B}_n\}$  be two sequences of  $n \times n$  symmetric matrices. Assume the following limits exist:

$$\begin{aligned} w_1 &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr}(\mathbf{A}_n \odot \mathbf{A}_n), & w_2 &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr}(\mathbf{B}_n \odot \mathbf{B}_n), \\ w_3 &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr}(\mathbf{A}_n \odot \mathbf{B}_n), \\ \theta_1 &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr}(\mathbf{A}_n \mathbf{A}_n^\top), & \theta_2 &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr}(\mathbf{B}_n \mathbf{B}_n^\top), \\ \theta_3 &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr}(\mathbf{A}_n \mathbf{B}_n^\top), \end{aligned}$$

where  $\mathbf{A} \odot \mathbf{B}$  denotes the Hadamard product of two matrices  $\mathbf{A}$  and  $\mathbf{B}$ , that is,  $[\mathbf{A} \odot \mathbf{B}]_{ij} := [\mathbf{A}]_{ij} [\mathbf{B}]_{ij}$ . Define two groups of sesquilinear forms:

$$\begin{aligned} U(\ell) &= \frac{1}{\sqrt{n}} (X(\ell)^\top \mathbf{A}_n Y(\ell) - \rho(\ell) \text{Tr}(\mathbf{A}_n)), \\ V(\ell) &= \frac{1}{\sqrt{n}} (X(\ell)^\top \mathbf{B}_n Y(\ell) - \rho(\ell) \text{Tr}(\mathbf{B}_n)). \end{aligned}$$

Then the  $2K$ -dimensional random vector  $(U(1), \dots, U(K), V(1), \dots, V(K))^\top$  weakly converges to a zero-mean Gaussian vector with covariance matrix  $\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}_{2K \times 2K}$ , with each block  $\mathbf{B}_{ij} = (\mathbf{B}_{ij}(\ell, \ell'))_{1 \leq \ell, \ell' \leq K}$  a  $K \times K$  matrix having structure, for  $1 \leq \ell, \ell' \leq K$ ,

$$\begin{aligned} \mathbf{B}_{11}(\ell, \ell') &= \text{Cov}(U(\ell), U(\ell')) = w_1 a_1 + (\theta_1 - w_1)(a_2 + a_3), \\ \mathbf{B}_{22}(\ell, \ell') &= \text{Cov}(V(\ell), V(\ell')) = w_2 a_1 + (\theta_2 - w_2)(a_2 + a_3), \\ \mathbf{B}_{12}(\ell, \ell') &= \text{Cov}(U(\ell), V(\ell')) = w_3 a_1 + (\theta_3 - w_3)(a_2 + a_3), \\ a_1 &= \mathbb{E}(X_{\ell 1} Y_{\ell 1} X_{\ell' 1} Y_{\ell' 1}) - \rho(\ell) \rho(\ell'), \\ a_2 &= \mathbb{E}(X_{\ell 1} X_{\ell' 1}) \mathbb{E}(Y_{\ell 1} Y_{\ell' 1}), & a_3 &= \mathbb{E}(X_{\ell 1} Y_{\ell' 1}) \mathbb{E}(X_{\ell' 1} Y_{\ell 1}). \end{aligned}$$

Noting that

$$\mathbf{R}_n(\ell) = \frac{1}{\sqrt{n}} \left( \mathbf{Y}_1 \left( \ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \mathbf{Y}_1^\top - \mathbf{I}_m \text{Tr} \left( \ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \right),$$

it can be seen that, for any  $1 \leq i, j, i', j' \leq m$ , the random vector  $([\mathbf{R}_n(\psi_{k_1})]_{ij}, [\mathbf{R}_n(\psi_{k_2})]_{i'j'})^\top$  forms a sesquilinear pair

$$\left( U(\ell) = \frac{1}{\sqrt{n}} (X(\ell)^\top \mathbf{A}_n Y(\ell) - \rho(\ell) \text{Tr}(\mathbf{A}_n)), V(\ell') = \frac{1}{\sqrt{n}} (X(\ell')^\top \mathbf{B}_n Y(\ell') - \rho(\ell') \text{Tr}(\mathbf{B}_n)) \right)^\top,$$

where

$$\begin{aligned} A_n &= \left( \psi_{k_1} \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1}, & B_n &= \left( \psi_{k_2} \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1}, \\ \rho(\ell) &= \mathbb{E}(X_{\ell 1} Y_{\ell 1}) = \delta_{ij}, & \rho(\ell') &= \delta_{i'j'}, \end{aligned}$$

and  $X(\ell)$  corresponds to the  $i$ th row of  $\mathbf{Y}_1$ ,  $Y(\ell)$  corresponds to the  $j$ th row of  $\mathbf{Y}_1$ ,  $X(\ell')$  corresponds to the  $i'$ th row of  $\mathbf{Y}_1$ ,  $Y(\ell')$  corresponds to the  $j'$ th row of  $\mathbf{Y}_1$ .

Therefore,

$$w_3 = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr} \left( \left( \psi_{k_1} \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \odot \left( \psi_{k_2} \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \right) = \underline{s}(\psi_{k_1}) \underline{s}(\psi_{k_2}),$$

$$\theta_3 = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr} \left( \left( \psi_{k_1} \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'ii} \mathbf{Y}_2 \right)^{-1} \left( \psi_{k_2} \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \right) = \frac{\underline{s}(\psi_{k_1}) - \underline{s}(\psi_{k_2})}{\psi_{k_1} - \psi_{k_2}}.$$

By Lemma 5.3, for any  $1 \leq i, j, i', j' \leq m$ ,  $([\mathbf{R}_n(\psi_{k_1})]_{ij}, [\mathbf{R}_n(\psi_{k_2})]_{i'j'})^\top$  weakly converges to a zero-mean Gaussian vector  $([\mathbf{R}(\psi_{k_1})]_{ij}, [\mathbf{R}(\psi_{k_2})]_{i'j'})^\top$  with the following covariance structure:

(1) for  $i = i', j = j', i \neq j, a_1 = a_2 = 1, a_3 = 0$ ,

$$\text{Cov}([\mathbf{R}(\psi_{k_1})]_{ij}, [\mathbf{R}(\psi_{k_2})]_{ij}) = \theta_3 = \frac{\underline{s}(\psi_{k_1}) - \underline{s}(\psi_{k_2})}{\psi_{k_1} - \psi_{k_2}};$$

(2) for  $i = i' = j = j', a_1 = \nu_4 - 1, a_2 = 1, a_3 = 1$ ,

$$\text{Cov}([\mathbf{R}(\psi_{k_1})]_{ii}, [\mathbf{R}(\psi_{k_2})]_{ii}) = 2\theta_3 + (\nu_4 - 3)w_3 = \frac{2(\underline{s}(\psi_{k_1}) - \underline{s}(\psi_{k_2}))}{\psi_{k_1} - \psi_{k_2}} + \beta_y \underline{s}(\psi_{k_1}) \underline{s}(\psi_{k_2});$$

(3) for all the other cases,  $a_1 = a_2 = a_3 = 0$ ,

$$\text{Cov}([\mathbf{R}(\psi_{k_1})]_{ij}, [\mathbf{R}(\psi_{k_2})]_{i'j'}) = 0.$$

Then substituting  $\underline{s}(\psi_k)$  with (5.5), the limiting distribution of  $(\sqrt{n}(\frac{\lambda_1}{\psi_1} - 1), \dots, \sqrt{n}(\frac{\lambda_m}{\psi_m} - 1))^\top$  in Theorem 3.1 naturally follows.

*Step 2: Marginal limiting distribution of  $\text{Tr}(\mathbf{S}_{22})$  and  $\text{Tr}(\mathbf{S}_{11})$ .*

In this step, we study the marginal limiting distribution of  $\text{Tr}(\mathbf{S}_{22})$ . In fact,

$$\text{Tr}(\mathbf{S}_{22}) = \text{Tr} \left( \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right) = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i^\top \mathbf{V}_{p'} \mathbf{Y}_i,$$

where each  $\mathbf{Y}_i$  is a random vector with  $p'$  i.i.d. entries  $Y_{ij} (1 \leq j \leq p')$  satisfying  $\mathbb{E}Y_{ij} = 0, \mathbb{E}Y_{ij}^2 = 1, \mathbb{E}Y_{ij}^4 = \nu_4$ .

Moreover, by some calculations, we have

$$\mathbb{E}(\text{Tr}(\mathbf{S}_{22})) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\mathbf{Y}_i^\top \mathbf{V}_{p'} \mathbf{Y}_i) = \text{Tr}(\mathbf{V}_{p'}),$$

$$\begin{aligned} \mathbb{E}(\text{Tr}(\mathbf{S}_{22}))^2 &= \frac{1}{n^2} \sum_{i \neq j}^n \mathbb{E}(\mathbf{Y}_i^\top \mathbf{V}_{p'} \mathbf{Y}_i \mathbf{Y}_j^\top \mathbf{V}_{p'} \mathbf{Y}_j) + \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}(\mathbf{Y}_i^\top \mathbf{V}_{p'} \mathbf{Y}_i)^2 \\ &= \frac{n^2 - n}{n^2} \text{Tr}^2(\mathbf{V}_{p'}) + \frac{1}{n} \left[ \text{Tr}^2(\mathbf{V}_{p'}) + 2 \text{Tr}(\mathbf{V}_{p'}^2) + (\nu_4 - 3) \sum_{i=1}^{p'} [\mathbf{V}_{p'}]_{ii}^2 \right]. \end{aligned}$$

Thus

$$\text{Var}(\text{Tr}(\mathbf{S}_{22})) = \frac{2}{n} \text{Tr}(\mathbf{V}_{p'}^2) + \frac{\nu_4 - 3}{n} \sum_{i=1}^{p'} [\mathbf{V}_{p'}]_{ii}^2.$$

Actually, in Section 4.2.3 of [7], the authors have proved the asymptotic normality for trace of any symmetric polynomial of a general class of sample (auto)covariance matrices. It is

directly applicable to our case of  $\text{Tr}(\mathbf{S}_{22})$  since our model settings fulfill all their assumptions. Therefore, we have, as  $p, n \rightarrow \infty, p/n \rightarrow y > 0$ ,

$$\text{Tr}(\mathbf{S}_{22}) - \text{Tr}(\mathbf{V}_{p'}) \xrightarrow{d} \mathcal{N}(0, 2y\gamma_2 + y(\nu_4 - 3)\gamma_{d,2}),$$

where  $\gamma_2 = \lim_{p' \rightarrow \infty} \frac{1}{p'} \text{Tr}(\mathbf{V}_{p'}^2)$ ,  $\gamma_{d,2} = \lim_{p' \rightarrow \infty} \frac{1}{p'} \sum_{i=1}^{p'} [\mathbf{V}_{p'}]_{ii}^2$ .

Similarly, we have

$$\mathbb{E}(\text{Tr}(\mathbf{S}_{11})) = \text{Tr}(\mathbf{\Lambda}), \quad \text{Var}(\text{Tr}(\mathbf{S}_{11})) = \frac{2}{n} \text{Tr}(\mathbf{\Lambda}^2) + \frac{\nu_4 - 3}{n} \sum_{i=1}^m [\mathbf{\Lambda}]_{ii}^2.$$

By Linderberg–Feller central limit theorem, as  $n \rightarrow \infty$ ,

$$(5.8) \quad \sqrt{n}(\text{Tr}(\mathbf{S}_{11}) - \text{Tr}(\mathbf{\Lambda})) \xrightarrow{d} \mathcal{N}\left(0, \sum_{i=1}^m \Lambda_{ii}^2(\nu_4 - 1) + \sum_{i \neq j} \Lambda_{ij}^2\right).$$

*Step 3: Joint limiting distribution of  $((\lambda_k)_{1 \leq k \leq m}, \text{Tr}(\mathbf{S}_n))^\top$ .*

First, by (5.8), we have

$$\text{Tr}(\mathbf{S}_n) - \text{Tr}(\mathbf{\Sigma}_p) = \text{Tr}(\mathbf{S}_{11}) - \text{Tr} \mathbf{\Lambda} + \text{Tr}(\mathbf{S}_{22}) - \text{Tr}(\mathbf{V}_{p'}) = \text{Tr}(\mathbf{S}_{22}) - \text{Tr}(\mathbf{V}_{p'}) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

Thus  $\text{Tr}(\mathbf{S}_n) - \text{Tr}(\mathbf{\Sigma}_p)$  shares the same Gaussian limiting distribution with  $\text{Tr}(\mathbf{S}_{22}) - \text{Tr}(\mathbf{V}_{p'})$ , that is, under Assumptions (i) to (iv),

$$\text{Tr}(\mathbf{S}_n) - \text{Tr}(\mathbf{\Sigma}_p) \xrightarrow{d} \mathcal{N}(0, 2y\gamma_2 + y(\nu_4 - 3)\gamma_{d,2}).$$

Second, from the previous proof we know that the main fluctuation of  $\lambda'_k$ s originates from  $\mathbf{R}_n(\ell)$ . It can be seen that  $(\lambda_k)_{1 \leq k \leq m}$  are asymptotically independent of  $\mathbf{Y}_2$  because  $\mathbf{R}_n(\ell)$  weakly converges to a Gaussian matrix with distribution independent of  $\mathbf{Y}_2$ . Actually, by going through the proof of Theorem 11.10 (see Lemma 5.2 in this paper) and the result of Theorem 10.8 in [29], it can be proved that, conditioning on  $\mathbf{Y}_2$ , the limiting distribution of  $\mathbf{R}_n(\ell)$  is a function of the LSD of  $\mathbf{S}_{22}$ , which does not depend on the value of the conditioning variable  $\mathbf{Y}_2$ .

This establishes the asymptotic independence between  $\mathbf{R}_n(\ell)$  and  $\mathbf{Y}_2$ . Moreover, since

$$\mathbf{R}_n(\ell) = \frac{1}{\sqrt{n}} \left( \mathbf{Y}_1 \left( \ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \mathbf{Y}_1^\top - \mathbf{I}_m \text{Tr} \left( \ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} \right),$$

if we treat  $\frac{1}{\sqrt{n}} \mathbf{Y}_1 (\ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2)^{-1} \mathbf{Y}_1^\top$  as  $f(\mathbf{Y}_1, \mathbf{Y}_2)$ , then

$$\begin{aligned} \frac{1}{\sqrt{n}} \mathbf{I}_m \text{Tr} \left( \ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1} &= \mathbb{E}(f(\mathbf{Y}_1, \mathbf{Y}_2) | \mathbf{Y}_2), \\ \mathbf{R}_n(\ell) &= f(\mathbf{Y}_1, \mathbf{Y}_2) - \mathbb{E}(f(\mathbf{Y}_1, \mathbf{Y}_2) | \mathbf{Y}_2), \quad \text{Cov}(\mathbf{R}_n(\ell), \mathbf{Y}_2) = \mathbf{0}. \end{aligned}$$

Note that  $\text{Tr}(\mathbf{S}_{22}) = \frac{1}{n} \text{Tr}(\mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2)$ , the randomness of  $\text{Tr}(\mathbf{S}_{22})$  all originates from  $\mathbf{Y}_2$  and marginally  $\text{Tr}(\mathbf{S}_{22})$  is also asymptotically normal. Accordingly, we have  $((\lambda_k)_{1 \leq k \leq m}, \text{Tr}(\mathbf{S}_{22}))^\top$  are asymptotically normal and independent.

Since  $\text{Tr}(\mathbf{S}_{11})$  is of order  $O(1/\sqrt{n})$  and  $(\lambda_k)_{1 \leq k \leq m}$  is of constant order,  $((\lambda_k)_{1 \leq k \leq m}, \text{Tr}(\mathbf{S}_n))^\top$  is also asymptotically normally distributed and the covariance between  $(\lambda_k)_{1 \leq k \leq m}$  and  $\text{Tr}(\mathbf{S}_n)$  is of order  $O(1/\sqrt{n})$ .

5.2. *Asymptotic joint distribution of  $((\lambda_k)_{1 \leq k \leq m}, \text{Tr}(\mathbf{S}_{11}))^\top$ .* In this section, we present a result of the joint distribution of  $((\lambda_k)_{1 \leq k \leq m}, \text{Tr}(\mathbf{S}_{11}))^\top$ . It is crucial for quantifying second-order terms in Theorem 3.2.

Without loss of generality, we consider  $(\lambda_k, \text{Tr}(\mathbf{S}_{11}))^\top$  first. Since

$$\text{Tr}(\mathbf{S}_{11}) - \text{Tr}(\mathbf{\Lambda}) = \frac{1}{n} \text{Tr}(\mathbf{Y}_1^\top \mathbf{\Lambda} \mathbf{Y}_1) - \text{Tr}(\mathbf{\Lambda}),$$

the leading term of  $\delta_{n,k} = \sqrt{n}(\frac{\lambda_k}{\psi_k} - 1)$  is

$$\frac{[\mathbf{U}^\top \mathbf{R}_n(\psi_k) \mathbf{U}]_{kk}}{\underline{s}'(\psi_k)\psi_k} + \frac{\sqrt{n}(\underline{s}(\psi_k) - \underline{s}_n(\psi_k))}{\psi_k \underline{s}'(\psi_k)}.$$

The second term  $\underline{s}_n(\psi_k)$  is a function of  $\mathbf{Y}_2$  which is independent from  $\text{Tr}(\mathbf{S}_{11})$ . Therefore, we only have to consider the correlation between  $\mathbf{R}_n(\psi_k)$  and  $\text{Tr}(\mathbf{S}_{11})$ .

Note that  $\text{Tr}(\mathbf{S}_{11}) = \frac{1}{n} \text{Tr}(\mathbf{\Lambda} \mathbf{Y}_1 \mathbf{Y}_1^\top)$  can be seen as linear combinations of entries in the  $m \times m$  matrix  $\frac{1}{n} \mathbf{Y}_1 \mathbf{Y}_1^\top$ . According to Lemma 5.3,

$(\mathbf{R}_n(\ell), \frac{1}{\sqrt{n}}(\mathbf{Y}_1 \mathbf{Y}_1^\top - \mathbf{I}_m))^\top$  forms a random sesquilinear pair with

$$\mathbf{A}_n = \left( \ell \mathbf{I}_n - \frac{1}{n} \mathbf{Y}_2^\top \mathbf{V}_{p'} \mathbf{Y}_2 \right)^{-1}, \quad \mathbf{B}_n = \mathbf{I}_n.$$

If the correlation between each entry of  $\mathbf{R}_n(\ell)$  and  $\frac{1}{\sqrt{n}}(\mathbf{Y}_1 \mathbf{Y}_1^\top - \mathbf{I}_m)$  can be obtained, then we can derive the joint distribution of  $(\lambda_k, \text{Tr}(\mathbf{S}_{11}))^\top$ . More specifically, we have the following result, whose proof is relegated to the Supplementary Material [19].

PROPOSITION 5.1. *Under Assumptions (i)–(iv), as  $p, n \rightarrow \infty, p/n \rightarrow y$ , we have*

$$\left( \begin{array}{c} \sqrt{n} \left( \frac{\lambda_k}{\psi_k} - 1 \right) \\ \sqrt{n} (\text{Tr}(\mathbf{S}_{11}) - \text{Tr}(\mathbf{\Lambda})) \end{array} \right) \xrightarrow{d} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho_k \\ \rho_k & \sigma_2^2 \end{pmatrix} \right),$$

where

$$\sigma_1^2 = \sum_{i=1}^m u_{ik}^4 \sigma_{\alpha_k}^2 + \sum_{i \neq j}^m u_{ik}^2 u_{jk}^2 s_{\alpha_k}^2, \quad \sigma_2^2 = \sum_{i=1}^m \Lambda_{ii}^2 (v_4 - 1) + \sum_{i \neq j}^m \Lambda_{ij}^2,$$

$$\rho_k = \frac{\alpha_k \psi'(\alpha_k)}{\psi_k} \left( (v_4 - 1) \sum_{i=1}^m \Lambda_{ii} u_{ik}^2 + \sum_{i \neq j}^m \Lambda_{ij} u_{ik} u_{jk} \right),$$

$\mathbf{\Lambda} = (\Lambda_{ij})_{m \times m}$ , has spectral decomposition  $\mathbf{\Lambda} = \mathbf{U} \text{diag}(\alpha_1, \dots, \alpha_m) \mathbf{U}^\top$ ,  $\mathbf{U} = (u_{ij})_{m \times m}$ ,  $\sigma_{\alpha_k}^2$  and  $s_{\alpha_k}^2$  are defined in (3.1) and (3.2).

**Acknowledgments.** The authors thank the Co-editor, Richard Samworth, the Associate Editor and two referees for many constructive and motivating comments which greatly improve the paper. The research of F. Han was supported in part by NSF Grant DMS-1712536. J. Yao thanks support from Hong Kong RGC grant GRF-17306918. The research of Z. Li was supported by NSFC (National Natural Science Foundation of China) Grant No. 12031005.

SUPPLEMENTARY MATERIAL

Supplement to “Asymptotic joint distribution of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model” (DOI: 10.1214/19-AOS1882SUPP; .pdf). This supplement contains the proofs of Theorem 3.2 and Proposition 5.1 and extra simulation results.

## REFERENCES

- [1] BAI, Z. and DING, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices Theory Appl.* **1** 1150011, 21. MR2934717 <https://doi.org/10.1142/S2010326311500110>
- [2] BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053 <https://doi.org/10.1214/07-AIHP118>
- [3] BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686 <https://doi.org/10.1016/j.jmva.2011.10.009>
- [4] BAI, Z. and SILVERSTEIN, J. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 <https://doi.org/10.1214/aop/1078415845>
- [5] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 <https://doi.org/10.1214/009117905000000233>
- [6] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 <https://doi.org/10.1016/j.jmva.2005.08.003>
- [7] BHATTACHARJEE, M. and BOSE, A. (2016). Large sample behaviour of high dimensional autocovariance matrices. *Ann. Statist.* **44** 598–628. MR3476611 <https://doi.org/10.1214/15-AOS1378>
- [8] BIANCHI, P., DEBBAH, M., MAIDA, M. and NAJIM, J. (2011). Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inf. Theory* **57** 2400–2419. MR2809098 <https://doi.org/10.1109/TIT.2011.2111710>
- [9] CHEN, B. and PAN, G. (2015). CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli* **21** 1089–1133. MR3338658 <https://doi.org/10.3150/14-BEJ599>
- [10] CHOI, Y., TAYLOR, J. and TIBSHIRANI, R. (2017). Selecting the number of principal components: Estimation of the true rank of a noisy matrix. *Ann. Statist.* **45** 2590–2617. MR3737903 <https://doi.org/10.1214/16-AOS1536>
- [11] CHOW, T. L. and TEUGELS, J. L. (1979). The sum and the maximum of i.i.d. random variables. In *Proceedings of the Second Prague Symposium on Asymptotic Statistics (Hradec Králové, 1978)* 81–92. North-Holland, Amsterdam. MR0571177
- [12] DAVIS, R. A., HEINY, J., MIKOSCH, T. and XIE, X. (2016). Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes* **19** 517–547. MR3535965 <https://doi.org/10.1007/s10687-016-0251-7>
- [13] DEO, R. S. (2016). On the Tracy–Widom approximation of studentized extreme eigenvalues of Wishart matrices. *J. Multivariate Anal.* **147** 265–272. MR3484182 <https://doi.org/10.1016/j.jmva.2016.01.010>
- [14] HSING, T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *Ann. Probab.* **23** 938–947. MR1334178
- [15] JOHNSON, D. E. and GRAYBILL, F. A. (1972). An analysis of a two-way model with interaction and no replication. *J. Amer. Statist. Assoc.* **67** 862–868. MR0400566
- [16] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 <https://doi.org/10.1214/aos/1009210544>
- [17] KNOWLES, A. and YIN, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. MR3704770 <https://doi.org/10.1007/s00440-016-0730-4>
- [18] KRITCHMAN, S. and NADLER, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemom. Intell. Lab. Syst.* **94** 19–32.
- [19] LI, Z., HAN, F. and YAO, J. (2020). Supplement to “Asymptotic joint distribution of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model.” <https://doi.org/10.1214/19-AOS1882SUPP>
- [20] MA, Z. (2012). Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli* **18** 322–359. MR2888709 <https://doi.org/10.3150/10-BEJ334>
- [21] NADLER, B. (2011). On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix. *J. Multivariate Anal.* **102** 363–371. MR2739121 <https://doi.org/10.1016/j.jmva.2010.10.005>
- [22] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808 <https://doi.org/10.1214/13-AOS1100>
- [23] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865
- [24] PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 <https://doi.org/10.1016/j.jspi.2013.09.005>

- [25] SILVERSTEIN, J. W. and CHOI, S.-I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* **54** 295–309. MR1345541 <https://doi.org/10.1006/jmva.1995.1058>
- [26] WANG, Q., SILVERSTEIN, J. W. and YAO, J. (2014). A note on the CLT of the LSS for sample covariance matrix from a spiked population model. *J. Multivariate Anal.* **130** 194–207. MR3229532 <https://doi.org/10.1016/j.jmva.2014.04.021>
- [27] WANG, Q., SU, Z. and YAO, J. (2014). Joint CLT for several random sesquilinear forms with applications to large-dimensional spiked population models. *Electron. J. Probab.* **19** no. 103, 28. MR3275855 <https://doi.org/10.1214/ejp.v19-3339>
- [28] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 <https://doi.org/10.1214/16-AOS1487>
- [29] YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis. Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554 <https://doi.org/10.1017/CBO9781107588080>
- [30] ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional  $F$ -matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263 <https://doi.org/10.1214/11-AIHP414>
- [31] ZHENG, S., BAI, Z. and YAO, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190 <https://doi.org/10.1214/14-AOS1292>
- [32] ZHENG, S., BAI, Z., YAO, J. and ZHU, H. (2016). CLT for linear spectral statistics of large dimensional sample covariance matrices with dependent data. Preprint. Available at [arXiv:1708.03749](https://arxiv.org/abs/1708.03749).