Journal of Multivariate Analysis 184 (2021) 104742

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Limiting laws for extreme eigenvalues of large-dimensional spiked Fisher matrices with a divergent number of spikes

Junshan Xie [a,1], Yicheng Zeng [b,1], Lixing Zhu [c,b,*]

[a] School of Mathematics and Statistics, Henan University, Kaifeng, China
[b] Department of Mathematics, Hong Kong Baptist University, Hong Kong
[c] Research Center for Statistics and Data Science, Beijing Normal University, Zhuhai, China

ARTICLE INFO

Article history:
Received 24 August 2019
Received in revised form 21 February 2021
Accepted 22 February 2021
Available online 4 March 2021

AMS 2010 subject classifications:
primary 62H10
secondary 62H99

Keywords:
Extreme eigenvalue
Fisher matrix
Phase transition phenomenon
Random matrix theory
Spiked population model

ABSTRACT

Consider the $p \times p$ matrix that is the product of a population covariance matrix and the inverse of another population covariance matrix. Suppose that their difference has a divergent rank with respect to p, when two samples of sizes n and T from the two populations are available, we construct its corresponding sample version. In the high-dimensional regime where both n and T are proportional to p, we investigate the limiting laws for extreme (spiked) eigenvalues of the sample (spiked) Fisher matrix when the number of spikes is divergent and these spikes are unbounded. We derive the convergence in probability of these spiked eigenvalues after scaling, and the central limit theorem for normalized spiked eigenvalues.

1. Introduction

In the last few decades, with the remarkable development in storage devices and computing capability, the demand for processing complex-structured data increases dramatically. One of the features, which is also the challenge caused by these data sets, is their high dimensions. The difficulty is that the classical limit theory for multivariate statistical analysis fails to ensure reliable inference for high-dimensional data analysis. Classical limit theorems require "small p large n" to keep their validity, which conflicts with the situation "large p large n" in high-dimensional settings in the sense that $p/n \to c > 0$ as the corresponding asymptotic properties become rather different. To attack the relevant issues, random matrix theory (RMT) serves as a powerful tool in addressing statistical problems in high dimensions. The first research of random matrices in multivariate statistics was about the Wishart matrices in [18]. Abundant research has been conducted for various topics in this field during the past half-century, especially in recent years. In the area of RMT in statistics, we refer to monographs [2,19] for systematical studies and [12] for a comprehensive review.

A relevant topic in multivariate statistics is about testing the equality of two covariance matrices:

$$H_0: \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 \quad \text{versus} \quad H_1: \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 + \mathbf{\Delta}, \tag{1}$$

* Corresponding author at: Department of Mathematics, Hong Kong Baptist University, Hong Kong.
E-mail address: lzhu@hkbu.edu.hk (L. Zhu).

1 These authors contributed equally to this work.

https://doi.org/10.1016/j.jmva.2021.104742

where $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are two covariance matrices corresponding to two p-variate populations, and $\mathbf{\Delta}$ is a non-negative definite matrix with rank q. Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be the sample covariance matrices from these two populations, respectively. When $\mathbf{S}_2$ is invertible, the random matrix $\mathbf{F} = \mathbf{S}_2^{-1}\mathbf{S}_1$ is called a Fisher matrix.

The difference between the null hypothesis and the alternative hypothesis relies on those extreme eigenvalues of $\mathbf{F}$. Under the null hypothesis, $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, [16] established the well-known Wacheter distribution as the limiting spectral distribution (LSD) of $\mathbf{F}$. Some extensions were built later (see examples in [13–15]). Furthermore, [1] pointed out the fact that the largest eigenvalue of $\mathbf{F}$ converges in probability to the upper bound of the support of the LSD of $\mathbf{F}$. Under the alternative hypothesis, $\mathbf{F}$ is called a spiked Fisher matrix (see [17]), because $\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1$ has a spiked structure similar to that of a spiked population model proposed by [10]. More specifically, the matrix $\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1$ is assumed to have the spectrum

$$\mathrm{spec}(\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1) = \{\lambda_1, \ldots, \lambda_q, 1, \ldots, 1\}, \tag{2}$$

where $\lambda_1 \geq \cdots \geq \lambda_q > 1$. When the rank q of $\mathbf{\Delta}$ is finite, [6] showed the phase transition phenomenon of the extreme eigenvalues of $\mathbf{F}$ under Gaussian population assumption. That is, for $1 \leq i \leq q$, the ith largest eigenvalue of $\mathbf{F}$ will depart from the upper bound of the support of LSD of $\mathbf{F}$ if and only if λ_1 exceeds a certain phase transition point. [17] relaxed Gaussian assumption and established central limit theorems for outlier eigenvalues of $\mathbf{F}$.

We in this paper consider, as a reasonable extension in theory and applications, the case of divergent q with respect to the dimension p. We will investigate convergence in probability and central limit theorems for the spiked eigenvalues of the spiked Fisher matrices. We formulate our problem as follows.

Assume that

$$\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T) = (y_{ij})_{1\leq i\leq p, 1\leq j\leq T} \in \mathbb{R}^{p\times T}, \quad \mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n) = (z_{ij})_{1\leq i\leq p, 1\leq j\leq n} \in \mathbb{R}^{p\times n} \tag{3}$$

are two independent arrays of independent real-valued random variables with zero mean and unit variance. We consider two samples $\{\mathbf{\Sigma}_1^{1/2}\mathbf{y}_i\}_{1\leq i\leq T}$ and $\{\mathbf{\Sigma}_2^{1/2}\mathbf{z}_i\}_{1\leq i\leq n}$, then their corresponding sample covariance matrices can respectively be written as

$$\mathbf{S}_1 = \frac{1}{T}\sum_{i=1}^{T}\mathbf{\Sigma}_1^{\frac12}\mathbf{y}_i\mathbf{y}_i^\top\mathbf{\Sigma}_1^{\frac12} = \frac{1}{T}\mathbf{\Sigma}_1^{\frac12}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Sigma}_1^{\frac12}, \quad \mathbf{S}_2 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{\Sigma}_2^{\frac12}\mathbf{z}_i\mathbf{z}_i^\top\mathbf{\Sigma}_2^{\frac12} = \frac{1}{n}\mathbf{\Sigma}_2^{\frac12}\mathbf{Z}\mathbf{Z}^\top\mathbf{\Sigma}_2^{\frac12}.$$

Also, define the Fisher matrix $\mathbf{F} := \mathbf{S}_2^{-1}\mathbf{S}_1$ as a sample version of matrix $\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1$. We aim to investigate asymptotics of the eigenvalues of $\mathbf{F}$. As the eigenvalues of $\mathbf{F}$ remain invariant under the linear transformation

$$(\mathbf{S}_1, \mathbf{S}_2) \to \left(\mathbf{\Sigma}_2^{-\frac12}\mathbf{S}_1\mathbf{\Sigma}_2^{-\frac12}, \mathbf{\Sigma}_2^{-\frac12}\mathbf{S}_2\mathbf{\Sigma}_2^{-\frac12}\right), \tag{4}$$

we can assume $\mathbf{\Sigma}_2 = \mathbf{I}_p$ throughout this paper without loss of generality. Under the assumption (2), the eigenvalues of $\mathbf{\Sigma}_1$ are $\lambda_1 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_p = 1$. Recalling (1) that $\mathbf{\Sigma}_1$ is a rank q perturbation of $\mathbf{\Sigma}_2 = \mathbf{I}_p$, we simply assume

$$\mathbf{\Sigma}_1 = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-q} \end{pmatrix}. \tag{5}$$

For the sake of brevity and readability, we write the eigenvalues of $\mathbf{F}$ in a descending order $\hat\lambda_1 \geq \cdots \geq \hat\lambda_p$, simplifying the double subscripts as single ones. Note that $\hat\lambda_i$ depends on the sample size n.

We now summarize some related work and our contributions. When all the spiked eigenvalues λ_i, $i \in \{1, \ldots, q\}$, with fixed q, are bounded, there are some results in the literature, for instance, the almost sure convergence (strong consistency) and the central limit theorem (CLT) [17], and the asymptotic Tracy–Widom distribution for the largest non-spiked eigenvalue [7,8]. In this paper, we consider the case where the number of spiked eigenvalues $q = q(p) \to \infty$ as $p \to \infty$, and the spiked eigenvalues λ_i, $i \in \{1, \ldots, q\}$ diverge as $p \to \infty$. To the best of our knowledge, there is no relevant result in the literature. A relevant paper is [5] which studied spiked population models, where the asymptotics for spiked eigenvalues, including the convergence in probability (weak consistency), the CLT, and the Tracy–Widom law for the largest nonspiked eigenvalue were built under a general framework. Unlike the case of fixed q and bounded spikes λ_i, $i \in \{1, \ldots, q\}$, normalization for $\hat\lambda_i$, $i \in \{1, \ldots, q\}$ is needed for the divergent q case. We consider the convergence of the normalized eigenvalues $\hat\lambda_i/\lambda_i$ and the CLT of $(\hat\lambda_i - \theta_i)/\theta_i$, where θ_i is a centered parameter defined later.

A basic approach for proving the asymptotics about spiked eigenvalues is to analyze a $q \times q$ random matrix corresponding to the q spikes. When q is bounded, [17] derived the almost sure entrywise convergence of the $q \times q$ matrix (and hence the convergence in norm) and then the almost sure limits of q spiked eigenvalues. This argument fails to work in the divergent q case where the entrywise convergence does not imply the convergence in norm of a $q \times q$ matrix. Instead, we use the CLT for random sesquilinear forms in [3] to derive the convergence rate of each entry, and then use Chebyshev's inequality to put all entries together to derive the convergence rate of the matrix in ℓ_∞ norm. In this way, we achieve the convergence in probability as well as the CLT of spiked eigenvalues (after proper normalization). This approach is similar to that used in [5], some technical assumptions will be imposed similarly.

The remaining parts of the paper are organized as follows. Section 2 contains the main results, including the convergence in probability of $\hat\lambda_i/\lambda_i$ and central limit theorems of $(\hat\lambda_i - \theta_i)/\theta_i$, for spiked eigenvalues of the spiked Fisher matrix $\mathbf{F}$. Here, θ_i, $i \in \{1, \ldots, q\}$, is a sequence of centering parameters defined in this section. In Section 3, we present the proofs of our main results. The proofs of some technical lemmas are displayed in Section 4.

2. Main results

2.1. Notations and assumptions

Considering the linear transformation (4), we assume that $\mathbf{\Sigma}_2 = \mathbf{I}_p$ without loss of generality, and then $\mathbf{\Sigma}_1$ has the structure as shown in (5). Further, we decompose $\mathbf{\Sigma}_{11}$ in (5) as

$$\mathbf{\Sigma}_{11} = \mathbf{U}^\top \mathbf{\Lambda}_1 \mathbf{U}. \tag{6}$$

Here, $\mathbf{U} \equiv (\mathbf{u}_1, \dots, \mathbf{u}_q)^\top$ is a $q \times q$ orthogonal matrix and

$$\mathbf{\Lambda}_1 = \mathrm{diag}(\underbrace{\lambda_1, \dots, \lambda_{N_1}}_{n_1}, \dots, \underbrace{\lambda_{N_{\ell-1}+1}, \dots, \lambda_q}_{n_\ell}),$$

where $\lambda_1 = \cdots = \lambda_{N_1} > \cdots > \lambda_{N_{\ell-1}+1} = \cdots = \lambda_q$ and $N_i := \sum_{j=1}^{i} n_j$ for $1 \le i \le \ell$. In this case, $\mathbf{\Sigma}_1$ can be decomposed as

$$\mathbf{\Sigma}_1 = \begin{pmatrix} \mathbf{U}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-q} \end{pmatrix} \begin{pmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-q} \end{pmatrix} \begin{pmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-q} \end{pmatrix} =: \begin{pmatrix} \mathbf{U}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-q} \end{pmatrix} \Lambda \begin{pmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-q} \end{pmatrix}.$$

Give the decompositions of the sample covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ as follows. We first decompose the matrices $\mathbf{Y}$ and $\mathbf{Z}$ defined in (3) as $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ and $\mathbf{Z} = (\mathbf{Z}_1^\top, \mathbf{Z}_2^\top)^\top$, where $\mathbf{Y}_1, \mathbf{Z}_1 \in \mathbb{R}^{q\times n}$ and $\mathbf{Y}_2, \mathbf{Z}_2 \in \mathbb{R}^{(p-q)\times n}$. Let $\mathbf{X} := \mathbf{\Sigma}_1^{1/2}\mathbf{Y}$. Then we similarly write $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$, where $\mathbf{X}_1 = \mathbf{\Sigma}_{11}^{1/2}\mathbf{Y}_1 = \mathbf{U}^\top\mathbf{\Lambda}_1^{1/2}\mathbf{U}\mathbf{Y}_1 \in \mathbb{R}^{q\times T}$ and $\mathbf{X}_2 = \mathbf{Y}_2 \in \mathbb{R}^{(p-q)\times T}$. It follows that

$$\mathbf{S}_1 = \begin{pmatrix} \frac{1}{T}\mathbf{X}_1\mathbf{X}_1^\top & \frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top \\ \frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top & \frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top \end{pmatrix}, \quad \mathbf{S}_2 = \begin{pmatrix} \frac{1}{n}\mathbf{Z}_1\mathbf{Z}_1^\top & \frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top \\ \frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top & \frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top \end{pmatrix}. \tag{7}$$

For $\lambda \in \mathbb{R} \setminus \{0\}$, introduce

$$\mathbf{F}_0 = \left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\left(\frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top\right), \quad \mathbf{M}(\lambda) = \mathbf{I}_{p-q} - \frac{\mathbf{F}_0}{\lambda},$$

$$\widetilde{m}_\theta(z) = \frac{1}{p-q}\mathrm{tr}\left(z\mathbf{I}_{p-q} - \frac{\mathbf{F}_0}{\theta}\right)^{-1}, \ \theta \in \mathbb{R}, \ z \in \mathbb{C}^+. \tag{8}$$

Let $\mu_1 \ge \cdots \ge \mu_{p-q}$ be the eigenvalues of the Fisher matrix $\mathbf{F}_0$. The empirical spectral distribution (ESD) of $\mathbf{F}_0$ is defined as

$$F_n(x) = \frac{1}{p-q}\sum_{j=1}^{p-q}\mathbf{1}_{\{\mu_j \le x\}}, \ x \in \mathbb{R}.$$

By the result in [17], under the assumption of $p/n \to y \in (0,1)$ and $p/T \to c > 0$, almost surely, the empirical spectral distribution F_n weakly converges to the limiting spectral distribution $F_{c,y}$, whose Stieltjes transform $\mathcal{S}(z) = \int_{-\infty}^{\infty}(x-z)^{-1}dF_{c,y}(x)$ satisfies, for $z \notin [a,b]$

$$\mathcal{S}(z) = \frac{1-c}{zc} - \frac{c[z(1-y)+1-c] + 2zy - c\sqrt{[z(1-y)+1-c]^2 - 4z}}{2zc(c+zy)}, \tag{9}$$

where $a = (1-\sqrt{c+y-cy})^2(1-y)^{-2}$ and $b = (1+\sqrt{c+y-cy})^2(1-y)^{-2}$.

For any complex matrix A, we use $s_i(A)$ to denote the ith largest singular value, and $\|A\|$ to denote the largest singular value throughout this paper. Write $a_n = \mathrm{O}_{\mathrm{a.s.}}(b_n)$ if it almost surely holds that $a_n = \mathrm{O}(b_n)$. Constant C may take different value from each appearance in this paper.

The following assumptions are required.

Assumption 1. $y_p := p/n \to y \in (0,1)$, $\tilde{y}_p := (p-q)/n$; $c_p := p/T \to c > 0$, $\tilde{c}_p := (p-q)/T$; $q = q(n) \to \infty$ as $n \to \infty$ but $q = \mathrm{o}(n^{\frac{1}{6}})$.

Assumption 2. For any $1 \le i \le q$, λ_i satisfies $q^2/\lambda_i \to 0$ and either of the two conditions:
(a). $\lambda_i^{-1}\sum_{j=1}^q \lambda_j = \mathrm{o}(q^{-\frac{1}{2}}n^{\frac{1}{4}})$ and $\lambda_i^{-2}\sum_{j=1}^q \lambda_j = \mathrm{o}(q^{-1})$; (b). $\lambda_i \sum_{j=1}^q \lambda_j^{-1} = \mathrm{o}(q^{-\frac{1}{2}}n^{\frac{1}{4}})$.

Assumption 3. Random vectors in $\{\mathbf{y}_i : 1 \le i \le T\} \bigcup \{\mathbf{z}_i : 1 \le i \le n\}$ are independent identically distributed, $\mathrm{E}z_{ij} = 0$, $\mathrm{E}|z_{ij}|^2 = 1$ for $1 \le i \le p$ and $1 \le j \le n$, and $\sup_{1\le i\le p}\mathrm{E}|z_{ij}|^4 < \infty$.

Assumption 4. There exists some constant $C > 1$ such that $\lambda_{N_i}/\lambda_{N_{i+1}} \ge C$ for any $1 \le i \le \ell - 1$.

Assumption 5. $\{\lambda_i\}_{1\le i\le q}$ are of bounded multiplicities, i.e., $\sup_{1\le i\le \ell} n_i < \infty$.

2.2. Weak consistency

The weak consistency of $\hat{\lambda}_i$ is stated below. Due to the fact that λ_i may go to infinity with n, consider the limit in probability for the ratio $\hat{\lambda}_i/\lambda_i$, $1 \le i \le q$.

Theorem 2.1. *Assume that Assumptions 1–3 hold. Then for all $1 \le i \le q$,*

$$\frac{\hat{\lambda}_i}{\lambda_i} = \frac{1}{1-y} + \mathrm{O}\left(y_p - y\right) + \kappa q \mathrm{O}_\mathrm{p}\left(n^{-\frac{1}{2}} + \lambda_i^{-1}\right),$$

where $\kappa := \min\{\kappa_1, \kappa_2\}$ with $\kappa_1 := q + \lambda_i^{-1}\sum_{j=1}^{q}\lambda_j$ and $\kappa_2 := q + \lambda_i \sum_{j=1}^{q}\lambda_j^{-1}$.

Remark 2.1. Note that the limit of the ratio $\hat{\lambda}_i/\lambda_i$ is $1/(1-y) > 1$, for all $1 \le i \le q$. It is different from the corresponding limit in the spiked population model with divergent q, which is 1 (see Theorem 2.1 in [5]). Roughly speaking, when we take $y \to 0$ with $1/(1-y) \to 1$, asymptotically, a spiked Fisher matrix behaves similarly to the sample covariance matrix in a spiked population model.

Remark 2.2. In the case of fixed q and bounded spikes λ_i, $1 \le i \le q$, Theorem 3.1 in [17] shows that almost surely the spiked eigenvalue $\hat{\lambda}_i$ converges to $\lambda_i(\lambda_i + c - 1)(\lambda_i - \lambda_i y - 1)^{-1}$ that converges to $1/(1-y)$ as $\lambda_i \to \infty$. Thus, Theorem 2.1 indicates that in the divergent q case, the result coincides with the result in the fixed q case in [17].

Remark 2.3. We only study the case with unbounded spikes in Theorem 2.1, but actually it is readily extended to handle the case with both bounded and unbounded spikes. Consider the model

$$\Sigma_1 = \begin{pmatrix} \mathbf{U}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-q} \end{pmatrix} \Lambda \begin{pmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-q} \end{pmatrix},$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_q, \lambda_{q+1}, \dots, \lambda_{q+q_0}, 1, \dots, 1)$, $q = \mathrm{o}(n^{1/6})$ and q_0 is bounded. Assume that the spikes $\lambda_1 \ge \cdots \ge \lambda_q$ are unbounded as in Theorem 2.1 and $\lambda_{q+1} \ge \cdots \ge \lambda_{q+q_0}$ are bounded. For $q+1 \le i \le q+q_0$, by Theorem A.10 in [2], we have

$$\hat{\lambda}_i = s_i\left(\mathbf{S}_2^{-1}\mathbf{S}_1\right) \le s_i\left(\mathbf{S}_1\right) s_1\left(\mathbf{S}_2^{-1}\right) \le s_i\left(\Sigma_1\right) s_1\left(\frac{1}{T}\mathbf{Y}\mathbf{Y}^\top\right) s_1\left(\mathbf{S}_2^{-1}\right) < \infty$$

almost surely. Then it holds that

$$\det\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\mathbf{Z}_1^\top - \frac{1}{T}\mathbf{X}_1\mathbf{X}_1^\top\right) \ne 0.$$

Similarly as the decomposition in (19) which will be introduced in the proof of Theorem 2.1 later, we have

$$\det\left\{\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_2\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top\right) - \left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_2\mathbf{Z}_1^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top\right)\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\mathbf{Z}_1^\top - \frac{1}{T}\mathbf{X}_1\mathbf{X}_1^\top\right)^{-1}\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top\right)\right\} = 0. \tag{10}$$

In the same manner as used in the proof of Theorem 2.1, it can be checked that

$$\left|\left|\left|\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_2\mathbf{Z}_1^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top\right)\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\mathbf{Z}_1^\top - \frac{1}{T}\mathbf{X}_1\mathbf{X}_1^\top\right)^{-1}\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top\right)\right|\right|\right|_\infty = \mathrm{o}_\mathrm{p}(1).$$

Then the solution of (10) is close to that of the equation

$$\det\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_2\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top\right) = 0. \tag{11}$$

Note that the solution of (11) is an eigenvalue of the spiked Fisher matrix $(\mathbf{Z}_2\mathbf{Z}_2^\top/n)^{-1}(\mathbf{X}_2\mathbf{X}_2^\top/T)$ which has been well studied by [17]. Thus, the weak consistency for all outliers $\hat{\lambda}_i$, $1 \le i \le q+q_0$, could be achieved by combining Theorem 3.1 in [17] and Theorem 2.1. Such a kind of extension could also be considered for CLT in Theorem 2.2.

2.3. Central limit theorem

As λ_i, $1 \le i \le q$, goes to infinity, the consistency of $\hat{\lambda}_i/\lambda_i$ in Theorem 2.1 does not mean that $(1-y)\hat{\lambda}_i$ is a good estimator of λ_i. In this section, we establish the CLT for $\hat{\lambda}_i$ to provide further properties.

We first introduce a centered parameter for $\hat{\lambda}_i$. Let $\theta_i \in \mathbb{R}$, $1 \le i \le q$, satisfy

$$1-\frac{1}{n}\mathrm{E}\left[\mathrm{tr}\left\{\mathbf{M}^{-1}(\theta_i)\right\}\right]=\frac{\lambda_i}{\theta_i}\left(1+\frac{1}{T}\mathrm{E}\left[\mathrm{tr}\left\{\mathbf{M}^{-1}(\theta_i)\frac{\mathbf{F}_0}{\theta_i}\right\}\right]\right), \tag{12}$$

and define δ_i, for $1 \le i \le q$, as

$$\delta_i=\frac{\hat{\lambda}_i-\theta_i}{\theta_i}. \tag{13}$$

By Lemma 3.1, we have that, as $n \to \infty$,

$$\frac{1}{p-q}\mathrm{E}\left[\mathrm{tr}\left\{\mathbf{M}^{-1}(\theta_i)\right\}\right]=\mathrm{E}\left\{\widetilde{m}_{\theta_i}(1)\right\}\to 1,\quad \frac{1}{p-q}\mathrm{E}\left[\mathrm{tr}\left\{\mathbf{M}^{-1}(\theta_i)\frac{\mathbf{F}_0}{\theta_i}\right\}\right]\to 0.$$

It follows by (12) that

$$\frac{\lambda_i}{\theta_i}=\left(1-\frac{p-q}{n}\right)+\mathrm{o}(1)\to 1-y.$$

Since Eq. (12) for θ_i is hard to solve, an alternative definition for θ_i is proposed as follows. Recall the definition of $\widetilde{m}_\theta(z)$ in (8):

$$\widetilde{m}_\theta(z)=\frac{1}{p-q}\mathrm{tr}\left(z\mathbf{I}_{p-q}-\frac{\mathbf{F}_0}{\theta}\right)^{-1},\ \theta\in\mathbb{R},\ z\in\mathbb{C}^+.$$

Denoting $f_\theta(x)=\theta/(\theta-x)$ for any fixed $\theta \in \mathbb{R}$, we have

$$\widetilde{m}_\theta(1)=\frac{1}{p-q}\mathrm{tr}\left(\mathbf{I}_{p-q}-\frac{\mathbf{F}_0}{\theta}\right)^{-1}=\int_{-\infty}^{\infty}\frac{\theta}{\theta-x}dF_n(x)=:F_n(f_\theta),$$

where F_n denotes the ESD of the matrix $\mathbf{F}_0$. By CLT for linear spectral statistics (LSS) of Fisher matrices (see Theorem 3.10 in [19]), for any fixed θ,

$$p\{F_n(f_\theta)-F_{\tilde{c}_p,\tilde{y}_p}(f_\theta)\}$$

converges weakly to a Gaussian variable. It follows that

$$\begin{aligned}\widetilde{m}_\theta(1)=&F_{\tilde{c}_p,\tilde{y}_p}(f_\theta)+\mathrm{O_p}(n^{-1})=-\theta\widetilde{\mathcal{S}}(\theta)+\mathrm{O_p}(n^{-1})\\=&\frac{\tilde{c}_p-1}{\tilde{c}_p}+\frac{\tilde{c}_p\{\theta(1-\tilde{y}_p)+1-\tilde{c}_p\}+2\theta\tilde{y}_p-\tilde{c}_p\sqrt{\{\theta(1-\tilde{y}_p)+1-\tilde{c}_p\}^2-4\theta}}{2\tilde{c}_p(\tilde{c}_p+\theta\tilde{y}_p)}+\mathrm{O_p}(n^{-1}),\end{aligned}$$

where $\widetilde{\mathcal{S}}(\cdot)$ denotes the Stieltjes transform of $F_{\tilde{c}_p,\tilde{y}_p}$. This leads to

$$\mathrm{E}\{\widetilde{m}_\theta(1)\}=-\theta\widetilde{\mathcal{S}}(\theta)+\mathrm{O}(n^{-1}). \tag{14}$$

The definition of θ_i in (12) can be rewritten as

$$1-\tilde{y}_p\mathrm{E}\left\{\widetilde{m}_{\theta_i}(1)\right\}=\frac{\lambda_i}{\theta_i}\left[1-\tilde{c}_p+\tilde{c}_p\mathrm{E}\left\{\widetilde{m}_{\theta_i}(1)\right\}\right].$$

By using (14), we can further rewrite (12) as

$$1+\tilde{y}_p\theta_i\widetilde{\mathcal{S}}(\theta_i)+\mathrm{O}(n^{-1})=\frac{\lambda_i}{\theta_i}\left\{1-\tilde{c}_p-\tilde{c}_p\theta_i\widetilde{\mathcal{S}}(\theta_i)+\mathrm{O}(n^{-1})\right\}. \tag{15}$$

Thus, we give another definition of θ_i by the following equation:

$$1+\tilde{y}_p\theta_i\widetilde{\mathcal{S}}(\theta_i)=\frac{\lambda_i}{\theta_i}\left\{1-\tilde{c}_p-\tilde{c}_p\theta_i\widetilde{\mathcal{S}}(\theta_i)\right\}. \tag{16}$$

It is notable that the CLT below is also applicable to θ_i defined by (16). Comparing (15) with (16), we can see that (15) defines a θ_i that nears the one defined in (16). Note that the scale of δ_i is $n^{-1/2}$. These two θ_i's lead to two δ_i's that are close to each other, and their difference is at most $O(n^{-1})$ smaller than the scale $n^{-1/2}$ of δ_i. Even Taylor's expansion on the Stieltjes transformation $\widetilde{\mathcal{S}}(\cdot)$ can be used to (16) to get an explicit form of θ_i although some errors would appear. In the remaining parts of this paper, we use θ_i defined by (12) in all results and their proofs.

Consider the case where all the spiked eigenvalues are simple, that is, $n_i = 1$ for all $1 \le i \le \ell$, which means that $\mathbf{\Lambda}_1 = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_q)$.

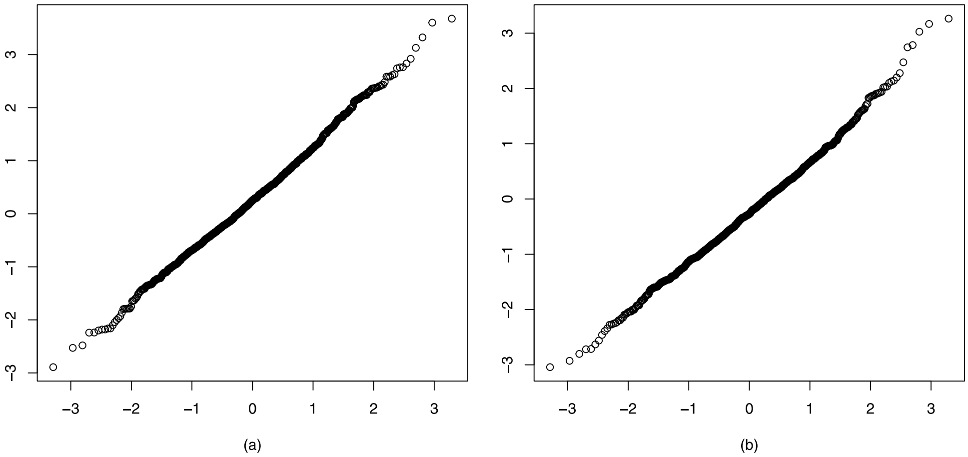

Fig. 1. (a) The qq plot of the normalized largest spiked eigenvalue $\sqrt{p}\delta_1/\sigma_1$ from 1000 independent replications. (b) The qq plot of the normalized smallest spiked eigenvalue $\sqrt{p}\delta_q/\sigma_q$ from 1000 independent replications.

Theorem 2.2. *Under Assumptions 1–4 and that $n_i = 1,\ 1 \le i \le \ell$, i.e., $\ell = q$, it holds that, for all $1 \le i \le q$,*

$$\sqrt{p}\frac{\delta_i}{\sigma_i} \xrightarrow{d} \mathcal{N}(0, 1)$$

with δ_i defined in (13) and $\sigma_i^2 := (y + c)\nu_i - c - y(1 - 3y)(1 - y)^{-1}$, where $\nu_i = \mathrm{E}|\mathbf{u}_i^\top \mathbf{Z}_1\mathbf{e}_1|^4$, $\mathbf{e}_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^q$ and $\mathbf{u}_i \in \mathbb{R}^q$ is the ith column of the matrix $\mathbf{U}^\top$ defined in (6).

Remark 2.4. When the variance σ_i^2 at the population level is unknown, an estimation is in need. A natural way would be to estimate the eigenvector $\mathbf{u}_i$ first. For the spiked population model, [5] showed that when a leading eigenvalue of the sample covariance matrix is divergent, the corresponding eigenvector is a good estimator for its population counterpart in terms of inner product. However, the situation becomes much more difficult when it comes to the spiked Fisher matrix. Recalling the assumed structure $\mathbf{\Sigma}_2^{-1/2}\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1/2} = \mathbf{I}_p + \mathbf{\Delta}$, we suppose that $\mathbf{v}_i := (\mathbf{u}_i^\top, 0, \ldots, 0)^\top \in \mathbb{R}^p$ is the eigenvector of $\mathbf{\Sigma}_2^{-1/2}\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1/2} = \mathbf{I}_p + \mathbf{\Delta}$ corresponding to λ_i and $\hat{\mathbf{v}}_i$ is that of $\mathbf{S}_1 = \mathbf{\Sigma}_1^{1/2}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Sigma}_1^{1/2}$. Then $\mathbf{\Sigma}_2^{1/2}\hat{\mathbf{v}}_i$ is the eigenvector of $(\mathbf{I}_p + \mathbf{\Delta})^{1/2}\mathbf{Y}\mathbf{Y}^\top(\mathbf{I}_p + \mathbf{\Delta})^{1/2}$ corresponding to the ith largest eigenvalue. If $\mathbf{\Sigma}_2$ is known or can be consistently estimated, $\mathbf{\Sigma}_2^{1/2}\hat{\mathbf{v}}_i$ is a good estimator of $\mathbf{v}_i$, by Theorem 4.1 in [5]. But actually $\mathbf{\Sigma}_2$ cannot be easily recovered based on $\mathbf{S}_2$ because of the delocalization of eigenvectors for non-outliers (see [4]). Thus, how to construct a consistent estimation of $\mathbf{\Sigma}_2$ becomes a challenging issue. As a special case, when entries of $\mathbf{Y}$ and $\mathbf{Z}$ are Gaussian, the parameter ν_i equals 3, which is independent of the value of $\mathbf{u}_i$. In practice, the bootstrap approximation would be an alternative way to achieve a reliable estimation of σ_i^2. For estimation of the variance of the largest sample eigenvalue in a spiked population model, [11] showed that the bootstrap approximation works when the largest eigenvalue is quite large. This deserves a further study.

To check the practical applicability of Theorem 2.2, a simulation is conducted. Set $p = 200$, $T = 600$, $n = 1000$, $q = \lceil 2\log p \rceil$, $\lambda_i = (3/2)^{q+1-i}(\log p/3)^3$ for $1 \le i \le q$, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. Let $\mathbf{\Sigma}_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_q, 1, \ldots, 1)$ and $\mathbf{\Sigma}_2 = \mathbf{I}_p$. Draw a sample $\{\mathbf{x}_i\}_{1\le i\le T}$ of size T from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_1)$ and a sample $\{\mathbf{z}_i\}_{1\le i\le n}$ of size n from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_2)$. Compute the largest q eigenvalues $\hat{\lambda}_i$, $1 \le i \le q$, of the Fisher matrix $\mathbf{F} = \mathbf{S}_2^{-1}\mathbf{S}_1$ and then δ_i accordingly, where $\mathbf{S}_1 = \sum_{i=1}^T \mathbf{x}_i\mathbf{x}_i^\top/T$ and $\mathbf{S}_2 = \sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i^\top/n$. We draw qq plots in Fig. 1 of $\sqrt{p}\delta_1/\sigma_1$ and $\sqrt{p}\delta_q/\sigma_q$ from 1000 independent replications. It suggests that both of $\sqrt{p}\delta_1/\sigma_1$ and $\sqrt{p}\delta_q/\sigma_q$ are approximated by the standard normal distribution.

Next, consider the case where some spiked eigenvalues are possibly multiple:

$$\mathbf{\Lambda}_1 = \mathrm{diag}(\underbrace{\lambda_1, \ldots, \lambda_{N_1}}_{n_1}, \ldots, \underbrace{\lambda_{N_{\ell-1}+1}, \ldots, \lambda_q}_{n_\ell}),$$

where $\lambda_1 = \cdots = \lambda_{N_1} > \cdots > \lambda_{N_{\ell-1}+1} = \cdots = \lambda_q$, $N_i := \sum_{j=1}^i n_j$ for $1 \le i \le \ell$ and there exists a constant $C < \infty$ such that $1 \le n_i \le C$ for all $1 \le i \le \ell$. According to the multiplicities of spiked eigenvalues, we divide the index set $\{1, \ldots, q\}$ into ℓ subsets, $J_i = \{N_{i-1} + 1, \ldots, N_i\}$, $1 \le i \le \ell$. Here we denote $N_0 = 0$. For any $1 \le i \le \ell$, and $1 \le h, k, h_1, k_1, h_2, k_2 \le n_i$, define

$$\mathcal{M}_{N_i,h,k} := \mathrm{E}\left(\mathbf{u}_{N_{i-1}+h}^\top \mathbf{Z}_1\mathbf{e}_1\mathbf{u}_{N_{i-1}+k}^\top \mathbf{Z}_1\mathbf{e}_1\right), \quad \mathcal{M}_{N_i,h_1,k_1,h_2,k_2} := \mathrm{E}\left(\mathbf{u}_{N_{i-1}+h_1}^\top \mathbf{Z}_1\mathbf{e}_1\mathbf{u}_{N_{i-1}+k_1}^\top \mathbf{Z}_1\mathbf{e}_1\mathbf{u}_{N_{i-1}+h_2}^\top \mathbf{Z}_1\mathbf{e}_1\mathbf{u}_{N_{i-1}+k_2}^\top \mathbf{Z}_1\mathbf{e}_1\right).$$

Theorem 2.3. *Suppose that Assumptions 1–5 hold. Define $\phi_i(\hat{\lambda}_j) = (\hat{\lambda}_j - \theta_j)/\theta_j$, for $1 \le i \le \ell$ and $j \in J_i$. Then $\sqrt{p}\{\phi_i(\hat{\lambda}_j), j \in J_i\}$ converges weakly to the distribution of the eigenvalues of the $n_i \times n_i$ random matrix $\Re^{(i)}$, where $\Re^{(i)} = \left(R^{(i)}_{hk}\right)_{1\le h,k\le n_i}$ is a symmetric matrix with independent Gaussian entries with mean zero and covariance function*

$$\mathrm{cov}\left(R^{(i)}_{h_1,k_1}, R^{(i)}_{h_2,k_2}\right) = (1-y)^{-2}\left\{\omega\left(\mathcal{M}_{N_i,h_1,k_1,h_2,k_2} - \mathcal{M}_{N_i,h_1,k_1}\mathcal{M}_{N_i,h_2,k_2}\right) + (\beta-\omega)\left(\mathcal{M}_{N_i,h_1,k_2}\mathcal{M}_{N_i,h_2,k_1} + \mathcal{M}_{N_i,h_1,h_2}\mathcal{M}_{N_i,k_1,k_2}\right)\right\},$$

where $\omega = (y+c)(1-y)^2$ and $\beta = y(1-y) + c(1-y)^2$.

3. Proofs of the theorems

We begin with a sketch of the proofs. The proof of Theorem 2.1 proceeds in three steps. First, we prove that any spiked eigenvalue $\hat{\lambda}_i$, $1 \le i \le q$, solves Eq. (21) whose left-hand side is the determinant of a $q \times q$ matrix that can be decomposed into four terms, namely $\mathbf{U}\Xi_{\mathbf{A}}\mathbf{U}^\top$, $\mathbf{U}\Xi_{\mathbf{B}}\mathbf{U}^\top$, $\mathbf{U}\Xi_{\mathbf{C}}\mathbf{U}^\top$ and $\mathbf{U}\Xi_{\mathbf{D}}\mathbf{U}^\top$ defined below. Second, we derive the limits of the entries of these four matrices and their convergence rates in ℓ_∞ norm, where CLT for random sesquilinear forms in [3] and Chebyshev's inequality are repeatedly used. Third, using the eigenvalue perturbation theorems in (21), we estimate the fluctuation of the scaled eigenvalue $\hat{\lambda}_i/\lambda_i$ and reach the result. As for the proof of Theorem 2.2, we also work on Eq. (21) in three main steps. First, we rewrite the matrix in (21) as the sum of $\mathbf{U}\Theta_{1n}\mathbf{U}^\top$, $\mathbf{U}\delta_i\Theta_{2n}\mathbf{U}^\top$ and $\mathbf{U}\Theta_{3n}\mathbf{U}^\top$. See Eq. (46). Second, we prove CLT for each diagonal entry of $\mathbf{U}\Theta_{1n}\mathbf{U}^\top$ (Lemma 3.2) and estimate the ℓ_∞ norm of $\mathbf{U}\Theta_{1n}\mathbf{U}^\top$ (Lemma 3.3), $\mathbf{U}\Theta_{2n}\mathbf{U}^\top$ (Lemma 3.4) and $\mathbf{U}\Theta_{3n}\mathbf{U}^\top$. Third, we expand the determinant in (46) by Leibniz formula and then achieve CLT for δ_i.

Proof of Theorem 2.1. We first show that for $1 \le i \le q$, $\hat{\lambda}_i$ converges to infinity at the same order with λ_i almost surely, i.e., there exists some constant $C > 1$ such that $C^{-1} < \hat{\lambda}_i/\lambda_i < C$ almost surely.

For any $1 \le i \le q$, by Theorem A.10 in [2], we have that

$$\hat{\lambda}_i = s_i(\mathbf{S}_2^{-1}\mathbf{S}_1) \le s_i(\mathbf{S}_1)s_1(\mathbf{S}_2^{-1}) = s_i(\mathbf{S}_1)s_p^{-1}(\mathbf{S}_2) \quad \text{and} \quad s_i(\mathbf{S}_1) \le s_i(\mathbf{S}_2^{-1}\mathbf{S}_1)s_1(\mathbf{S}_2).$$

Noting a basic fact that $s_1(\mathbf{S}_2) \to (1+\sqrt{y})^2$ and $s_p(\mathbf{S}_2) \to (1-\sqrt{y})^2 > 0$ almost surely, we have $0 < C_1 < \hat{\lambda}_i/s_i(\mathbf{S}_1) \le C_2 < +\infty$ almost surely for some constants C_1 and C_2. Again, by Theorem A.10 in [2] and Weyl's inequality, we have

$$s_i(\mathbf{S}_1) \le s_i(\boldsymbol{\Sigma}_1)s_1\left(\frac{1}{T}\mathbf{Y}\mathbf{Y}^\top\right) = \lambda_i s_1\left(\frac{1}{T}\mathbf{Y}\mathbf{Y}^\top\right)$$

and

$$s_i(\mathbf{S}_1) = s_i\left(\frac{1}{T}\mathbf{Y}^\top\boldsymbol{\Sigma}_1\mathbf{Y}\right) = s_i\left(\frac{1}{T}\mathbf{Y}_1^\top\boldsymbol{\Sigma}_{11}\mathbf{Y}_1 + \frac{1}{T}\mathbf{Y}_2^\top\mathbf{Y}_2\right) \ge s_i\left(\frac{1}{T}\mathbf{Y}_1^\top\boldsymbol{\Sigma}_{11}\mathbf{Y}_1\right) \ge s_i(\boldsymbol{\Sigma}_{11})s_q\left(\frac{1}{T}\mathbf{Y}_1\mathbf{Y}_1^\top\right) = \lambda_i s_q\left(\frac{1}{T}\mathbf{Y}_1\mathbf{Y}_1^\top\right).$$

Due to the fact that

$$s_1\left(\frac{1}{T}\mathbf{Y}\mathbf{Y}^\top\right) \to (1+\sqrt{c})^2, \quad s_q\left(\frac{1}{T}\mathbf{Y}_1\mathbf{Y}_1^\top\right) \to 1$$

almost surely, we have $0 < C_3 < s_i(\mathbf{S}_1)/\lambda_i < C_4 < +\infty$ almost surely for some constants C_3 and C_4. Thus, we conclude that $C^{-1} < \hat{\lambda}_i/\lambda_i < C$ almost surely for some constant C.

For any $1 \le i \le q$, by the definition of $\hat{\lambda}_i$, it solves the equation $\det\left(\hat{\lambda}_i\mathbf{I} - \mathbf{S}_2^{-1}\mathbf{S}_1\right) = 0$, or equivalently,

$$\det\left(\hat{\lambda}_i\mathbf{S}_2 - \mathbf{S}_1\right) = 0. \tag{17}$$

By the decompositions of $\mathbf{S}_1$ and $\mathbf{S}_2$ in (7), Eq. (17) can be rewritten as

$$\det\begin{pmatrix} \frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\mathbf{Z}_1^\top - \frac{1}{T}\mathbf{X}_1\mathbf{X}_1^\top & \frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top \\ \frac{\hat{\lambda}_i}{n}\mathbf{Z}_2\mathbf{Z}_1^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top & \frac{\hat{\lambda}_i}{n}\mathbf{Z}_2\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top \end{pmatrix} = 0. \tag{18}$$

By the formula of the determinant of partitioned matrices, we know that $\det\begin{pmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{pmatrix} = \det(\mathbf{D})\det(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})$ when $\mathbf{D}$ is nonsingular. As for $1 \le i \le q$, $\hat{\lambda}_i$ is an outlier eigenvalue of $\mathbf{S}_2^{-1}\mathbf{S}_1$ because $\hat{\lambda}_i$ goes to infinity at the same order with λ_i almost surely, which means

$$\det\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_2\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top\right) \ne 0,$$

then it follows by (18) that

$$\det\left\{\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\mathbf{Z}_1^\top - \frac{1}{T}\mathbf{X}_1\mathbf{X}_1^\top\right) - \left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top\right)\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_2\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top\right)^{-1}\left(\frac{\hat{\lambda}_i}{n}\mathbf{Z}_2\mathbf{Z}_1^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top\right)\right\} = 0. \tag{19}$$

For $\lambda \in \mathbb{R}$, defining

$$\mathbf{A}(\lambda) = \mathbf{Z}_2^\top \mathbf{M}^{-1}(\lambda)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2,\ \ \mathbf{B}(\lambda) = \mathbf{X}_2^\top \mathbf{M}^{-1}(\lambda)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{\lambda T}\mathbf{X}_2,$$
$$\mathbf{C}(\lambda) = \mathbf{Z}_2^\top \mathbf{M}^{-1}(\lambda)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{\lambda T}\mathbf{X}_2,\ \ \mathbf{D}(\lambda) = \mathbf{X}_2^\top \mathbf{M}^{-1}(\lambda)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{\lambda n}\mathbf{Z}_2,$$

it holds that $\mathbf{A}(\lambda) = \mathbf{A}(\lambda)^\top$, $\mathbf{B}(\lambda) = \mathbf{B}(\lambda)^\top$ and $T\mathbf{C}(\lambda) = n\mathbf{D}(\lambda)^\top$. Then some elementary calculations lead to

$$\det\left[\hat{\lambda}_i\frac{\mathbf{Z}_1\left\{\mathbf{I}_n - \mathbf{A}(\hat{\lambda}_i)\right\}\mathbf{Z}_1^\top}{n} - \frac{\mathbf{X}_1\left\{\mathbf{I}_T + \mathbf{B}(\hat{\lambda}_i)\right\}\mathbf{X}_1^\top}{T} + \hat{\lambda}_i\frac{\mathbf{Z}_1\mathbf{C}(\hat{\lambda}_i)\mathbf{X}_1^\top}{n} + \hat{\lambda}_i\frac{\mathbf{X}_1\mathbf{D}(\hat{\lambda}_i)\mathbf{Z}_1^\top}{T}\right] = 0. \tag{20}$$

To ease the notation, we define

$$\Xi_{\mathbf{A}} := \hat{\lambda}_i\frac{\mathbf{Z}_1\left\{\mathbf{I}_n - \mathbf{A}(\hat{\lambda}_i)\right\}\mathbf{Z}_1^\top}{n},\ \ \Xi_{\mathbf{B}} := \frac{\mathbf{X}_1\left\{\mathbf{I}_T + \mathbf{B}(\hat{\lambda}_i)\right\}\mathbf{X}_1^\top}{T},\ \ \Xi_{\mathbf{C}} := \hat{\lambda}_i\frac{\mathbf{Z}_1\mathbf{C}(\hat{\lambda}_i)\mathbf{X}_1^\top}{n},\ \ \Xi_{\mathbf{D}} := \hat{\lambda}_i\frac{\mathbf{X}_1\mathbf{D}(\hat{\lambda}_i)\mathbf{Z}_1^\top}{T}.$$

Multiplying the matrix in (20) by $\mathbf{U}$ on the left-hand side and by $\mathbf{U}^\top$ on the right-hand side, we have

$$\det\left\{\mathbf{U}\left(\Xi_{\mathbf{A}} - \Xi_{\mathbf{B}} + \Xi_{\mathbf{C}} + \Xi_{\mathbf{D}}\right)\mathbf{U}^\top\right\} = 0. \tag{21}$$

Next, we analyze these four terms in the determinant. For the term $\mathbf{U}\Xi_A\mathbf{U}^\top$, we first consider the decomposition

$$\frac{1}{n}\mathbf{Z}_1\left\{\mathbf{I}_n - \mathbf{A}(\hat{\lambda}_i)\right\}\mathbf{Z}_1^\top = \frac{1}{n}\mathbf{Z}_1\left\{\mathbf{I}_n - \mathbf{A}(\lambda_i)\right\}\mathbf{Z}_1^\top + \frac{1}{n}\mathbf{Z}_1\left\{\mathbf{A}(\hat{\lambda}_i) - \mathbf{A}(\lambda_i)\right\}\mathbf{Z}_1^\top.$$

The following lemma introduces the asymptotics of $\widetilde{m}_\theta(1)$ we shall use.

Lemma 3.1. *Suppose that Assumptions 1 and 3 hold. For $\widetilde{m}_\theta(z)$ defined in (8) and any $\theta \to \infty$, we have $\widetilde{m}_\theta(1) - 1 = O_{\mathrm{a.s.}}(\theta^{-1})$.*

By Lemma 3.1, we have $\widetilde{m}_{\lambda_i}(1) - 1 = O_{\mathrm{a.s.}}(\lambda_i^{-1})$, which implies

$$\frac{1}{n}\operatorname{tr}\left\{\mathbf{I}_n - \mathbf{A}(\lambda_i)\right\} = 1 - \frac{p-q}{n}\widetilde{m}_{\lambda_i}(1) = 1 - y_p + \frac{q}{n} + O_{\mathrm{a.s.}}(\lambda_i^{-1}).$$

Note that $\mathrm{E}(\mathbf{Z}_1\mathbf{Z}_1^\top/n) = \mathbf{I}_q$ and $(\mathbf{X}_1, \mathbf{Z}_1)$ is independent of $(\mathbf{X}_2, \mathbf{Z}_2)$. Under Assumption 3, by using Theorem 7.2 of [3], we have that, for all $1 \le j \le q$,

$$\mathbf{e}_j^\top\left[\frac{1}{n}\mathbf{Z}_1\left\{\mathbf{I}_n - \mathbf{A}(\lambda_i)\right\}\mathbf{Z}_1^\top\right]\mathbf{e}_j - \left\{1 - \frac{p-q}{n}\widetilde{m}_{\lambda_i}(1)\right\} = O_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right) \tag{22}$$

and

$$\mathrm{E}\left(\mathbf{e}_j^\top\left[\frac{1}{n}\mathbf{Z}_1\left\{\mathbf{I}_n - \mathbf{A}(\lambda_i)\right\}\mathbf{Z}_1^\top\right]\mathbf{e}_j - \left\{1 - \frac{p-q}{n}\widetilde{m}_{\lambda_i}(1)\right\}\right)^2 = O\left(n^{-1}\right) \tag{23}$$

for all $1 \le j \le q$. For those off-diagonal elements, we have that, for any $1 \le j_1 \ne j_2 \le q$,

$$\mathbf{e}_{j_1}^\top\left[\frac{1}{n}\mathbf{Z}_1\left\{\mathbf{I}_n - \mathbf{A}(\lambda_i)\right\}\mathbf{Z}_1^\top\right]\mathbf{e}_{j_2} = O_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right) \tag{24}$$

and

$$\mathrm{E}\left(\mathbf{e}_{j_1}^\top\left[\frac{1}{n}\mathbf{Z}_1\left\{\mathbf{I}_n - \mathbf{A}(\lambda_i)\right\}\mathbf{Z}_1^\top\right]\mathbf{e}_{j_2}\right)^2 = O\left(n^{-1}\right), \tag{25}$$

which is implied by Theorem 7.1 and Corollary 7.1 in [3]. Also we can write

$$\begin{aligned}
\mathbf{A}(\lambda_i) - \mathbf{A}\left(\hat{\lambda}_i\right) &= \mathbf{Z}_2^\top\left\{\mathbf{M}^{-1}(\lambda_i) - \mathbf{M}^{-1}\left(\hat{\lambda}_i\right)\right\}\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2 \\
&= \mathbf{Z}_2^\top\mathbf{M}^{-1}(\lambda_i)\left\{\mathbf{M}\left(\hat{\lambda}_i\right) - \mathbf{M}(\lambda_i)\right\}\mathbf{M}^{-1}\left(\hat{\lambda}_i\right)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2 \\
&= \left(\lambda_i^{-1} - \hat{\lambda}_i^{-1}\right)\mathbf{Z}_2^\top\mathbf{M}^{-1}(\lambda_i)\mathbf{F}_0\mathbf{M}^{-1}\left(\hat{\lambda}_i\right)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2.
\end{aligned}$$

It can be bounded by

$$\left\|\mathbf{A}(\lambda_i)-\mathbf{A}\left(\hat{\lambda}_i\right)\right\|=\left\|\left(\lambda_i^{-1}-\hat{\lambda}_i^{-1}\right)\mathbf{Z}_2^\top\mathbf{M}^{-1}(\lambda_i)\mathbf{F}_0\mathbf{M}^{-1}\left(\hat{\lambda}_i\right)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\right\|$$
$$\leq\left|\lambda_i^{-1}-\hat{\lambda}_i^{-1}\right|\left\|\frac{1}{\sqrt{n}}\mathbf{Z}_2^\top\right\|\left\|\mathbf{M}^{-1}(\lambda_i)\right\|\left\|\mathbf{F}_0\right\|\left\|\mathbf{M}^{-1}\left(\hat{\lambda}_i\right)\right\|\left\|\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\right\|\left\|\frac{1}{\sqrt{n}}\mathbf{Z}_2\right\|=O(\lambda_i^{-1})$$

almost surely. It follows that, for any $1\leq j_1,j_2\leq q$,

$$\mathbf{e}_{j_1}^\top\left[\frac{1}{n}\mathbf{Z}_1\left\{\mathbf{A}(\hat{\lambda}_i)-\mathbf{A}(\lambda_i)\right\}\mathbf{Z}_1^\top\right]\mathbf{e}_{j_2}=O_{\text{a.s.}}(\lambda_i^{-1}).\tag{26}$$

Combining (22), (24)) and (26), we can get that, for any $1\leq j\leq q$,

$$\mathbf{e}_{j}^\top\left[\frac{1}{n}\mathbf{Z}_1\left\{\mathbf{I}_n-\mathbf{A}(\hat{\lambda}_i)\right\}\mathbf{Z}_1^\top\right]\mathbf{e}_{j}-\left(1-\frac{p-q}{n}\right)=O_{\text{p}}\left(n^{-\frac{1}{2}}\right)+O_{\text{a.s.}}(\lambda_i^{-1})$$

and that, for any $1\leq j_1\neq j_2\leq q$,

$$\mathbf{e}_{j_1}^\top\left[\frac{1}{n}\mathbf{Z}_1\left\{\mathbf{I}_n-\mathbf{A}(\hat{\lambda}_i)\right\}\mathbf{Z}_1^\top\right]\mathbf{e}_{j_2}=O_{\text{p}}\left(n^{-\frac{1}{2}}\right)+O_{\text{a.s.}}(\lambda_i^{-1}).$$

Replacing $\mathbf{Z}_1$ by $\mathbf{UZ}_1$, it is easy to check that all the above conclusions still hold:

$$\mathbf{e}_{j}^\top\left[\frac{1}{n}\mathbf{UZ}_1\left\{\mathbf{I}_n-\mathbf{A}(\hat{\lambda}_i)\right\}\mathbf{Z}_1^\top\mathbf{U}^\top\right]\mathbf{e}_{j}=1-\frac{p-q}{n}+O_{\text{p}}\left(n^{-\frac{1}{2}}\right)+O_{\text{a.s.}}(\lambda_i^{-1})\tag{27}$$

for all $1\leq j\leq q$, and

$$\mathbf{e}_{j_1}^\top\left[\frac{1}{n}\mathbf{UZ}_1\left\{\mathbf{I}_n-\mathbf{A}(\hat{\lambda}_i)\right\}\mathbf{Z}_1^\top\mathbf{U}^\top\right]\mathbf{e}_{j_2}=O_{\text{p}}\left(n^{-\frac{1}{2}}\right)+O_{\text{a.s.}}(\lambda_i^{-1})\tag{28}$$

for all $1\leq j_1\neq j_2\leq q$. By the definition of $\Xi_{\mathbf{A}}$ in (20), together with (27) and (28), we can see that, for all $1\leq j\leq q$,

$$\mathbf{e}_{j}^\top\mathbf{U}\Xi_{\mathbf{A}}\mathbf{U}^\top\mathbf{e}_{j}=\hat{\lambda}_i\left(1-\frac{p-q}{n}\right)+\lambda_iO_{\text{p}}\left(n^{-\frac{1}{2}}\right)+O_{\text{a.s.}}(1)\tag{29}$$

and that, for all $1\leq j_1\neq j_2\leq q$,

$$\mathbf{e}_{j_1}^\top\mathbf{U}\Xi_{\mathbf{A}}\mathbf{U}^\top\mathbf{e}_{j_2}=\lambda_iO_{\text{p}}\left(n^{-\frac{1}{2}}\right)+O_{\text{a.s.}}(1).\tag{30}$$

For the term $\mathbf{U}\Xi_{\mathbf{B}}\mathbf{U}^\top$, by the definition of $\mathbf{X}_1$, we can derive that

$$\mathbf{U}\Xi_{\mathbf{B}}\mathbf{U}^\top=\frac{1}{T}\Lambda_1^{\frac{1}{2}}\mathbf{UY}_1\left\{\mathbf{I}_T+\mathbf{B}(\hat{\lambda}_i)\right\}\mathbf{Y}_1^\top\mathbf{U}^\top\Lambda_1^{\frac{1}{2}}=\frac{1}{T}\Lambda_1^{\frac{1}{2}}\mathbf{UY}_1\left\{\mathbf{I}_T+\mathbf{B}(\lambda_i)\right\}\mathbf{Y}_1^\top\mathbf{U}^\top\Lambda_1^{\frac{1}{2}}+\frac{1}{T}\Lambda_1^{\frac{1}{2}}\mathbf{UY}_1\left\{\mathbf{B}(\hat{\lambda}_i)-\mathbf{B}(\lambda_i)\right\}\mathbf{Y}_1^\top\mathbf{U}^\top\Lambda_1^{\frac{1}{2}},$$

where

$$\frac{1}{T}\text{tr}\left\{\mathbf{I}_T+\mathbf{B}(\lambda_i)\right\}=\frac{1}{T}\text{tr}\left\{\mathbf{I}_T+\mathbf{X}_2^\top\mathbf{M}^{-1}(\lambda_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{\lambda_iT}\mathbf{X}_2\right\}$$
$$=1+\frac{1}{T}\text{tr}\left\{\mathbf{M}^{-1}(\lambda_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{\lambda_iT}\mathbf{X}_2\mathbf{X}_2^\top\right\}=1+\frac{1}{T}\text{tr}\left\{\left(\mathbf{I}_{p-q}-\frac{\mathbf{F}_0}{\lambda_i}\right)^{-1}\frac{\mathbf{F}_0}{\lambda_i}\right\}=1+\frac{p-q}{T}\left\{\widetilde{m}_{\lambda_i}(1)-1\right\}$$

and

$$\mathbf{B}(\hat{\lambda}_i)-\mathbf{B}(\lambda_i)=\mathbf{X}_2^\top\left\{\hat{\lambda}_i^{-1}\mathbf{M}^{-1}\left(\hat{\lambda}_i\right)-\lambda_i^{-1}\mathbf{M}^{-1}(\lambda_i)\right\}\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2$$
$$=\hat{\lambda}_i^{-1}\lambda_i^{-1}\mathbf{X}_2^\top\mathbf{M}^{-1}\left(\hat{\lambda}_i\right)\left\{\lambda_i\mathbf{M}(\lambda_i)-\hat{\lambda}_i\mathbf{M}\left(\hat{\lambda}_i\right)\right\}\mathbf{M}^{-1}(\lambda_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2$$
$$=\left(\hat{\lambda}_i^{-1}-\lambda_i^{-1}\right)\mathbf{X}_2^\top\mathbf{M}^{-1}\left(\hat{\lambda}_i\right)\mathbf{M}^{-1}(\lambda_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2.$$

The same arguments for deriving (29) and (30) lead to that, for all $1\leq j\leq q$,

$$\mathbf{e}_{j}^\top\mathbf{U}\Xi_{\mathbf{B}}\mathbf{U}^\top\mathbf{e}_{j}=\lambda_j+\lambda_jO_p\left(n^{-\frac{1}{2}}\right)+\lambda_jO_{\text{a.s.}}(\lambda_i^{-1})\tag{31}$$

and that, for $1 \le j_1, j_2 \le q$,

$$\mathbf{e}_{j_1}^\top \mathbf{U}\boldsymbol{\Xi}_{\mathbf{B}}\mathbf{U}^\top \mathbf{e}_{j_2} = \lambda_{j_1}^{\frac{1}{2}}\lambda_{j_2}^{\frac{1}{2}} O_p\left(n^{-\frac{1}{2}}\right) + \lambda_{j_1}^{\frac{1}{2}}\lambda_{j_2}^{\frac{1}{2}} O_{\text{a.s.}}(\lambda_i^{-1}) \tag{32}$$

for all $1 \le j_1 \ne j_2 \le q$.

For the term $\mathbf{U}(\boldsymbol{\Xi}_{\mathbf{C}} + \boldsymbol{\Xi}_{\mathbf{D}})\mathbf{U}^\top$, by using the fact that $\mathbf{Y}_1 = \mathbf{U}^\top \boldsymbol{\Lambda}_1^{-\frac{1}{2}}\mathbf{U}\mathbf{X}_1$, we have that

$$\begin{aligned}
\mathbf{U}(\boldsymbol{\Xi}_{\mathbf{C}} + \boldsymbol{\Xi}_{\mathbf{D}})\mathbf{U}^\top &= \mathbf{U}\left\{\hat{\lambda}_i \frac{\mathbf{Z}_1\mathbf{C}(\hat{\lambda}_i)\mathbf{X}_1^\top}{n} + \hat{\lambda}_i \frac{\mathbf{X}_1\mathbf{D}(\hat{\lambda}_i)\mathbf{Z}_1^\top}{T}\right\}\mathbf{U}^\top \\
&= \mathbf{U}\left\{\lambda_i \frac{\mathbf{Z}_1\mathbf{C}(\lambda_i)\mathbf{X}_1^\top}{n} + \lambda_i \frac{\mathbf{X}_1\mathbf{D}(\lambda_i)\mathbf{Z}_1^\top}{T}\right\}\mathbf{U}^\top + \mathbf{U}\mathbf{Z}_1\left\{\hat{\lambda}_i\mathbf{C}(\hat{\lambda}_i) - \lambda_i\mathbf{C}(\lambda_i)\right\}\frac{1}{n}\mathbf{Y}_1^\top\mathbf{U}^\top\boldsymbol{\Lambda}_1^{\frac{1}{2}} \\
&\quad + \boldsymbol{\Lambda}_1^{\frac{1}{2}}\mathbf{U}\mathbf{Y}_1\left\{\hat{\lambda}_i\mathbf{D}(\hat{\lambda}_i) - \lambda_i\mathbf{D}(\lambda_i)\right\}\frac{1}{T}\mathbf{Z}_1^\top\mathbf{U}^\top,
\end{aligned}$$

and that

$$\begin{aligned}
\mathbf{U}\left\{\lambda_i \frac{\mathbf{Z}_1\mathbf{C}(\lambda_i)\mathbf{X}_1^\top}{n} + \lambda_i \frac{\mathbf{X}_1\mathbf{D}(\lambda_i)\mathbf{Z}_1^\top}{T}\right\}\mathbf{U}^\top &= \mathbf{U}\begin{pmatrix}\mathbf{Z}_1 & \mathbf{X}_1\end{pmatrix}\begin{pmatrix}\mathbf{O} & \frac{\lambda_i\mathbf{C}(\lambda_i)}{n} \\ \frac{\lambda_i\mathbf{D}(\lambda_i)}{T} & \mathbf{O}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_1^\top \\ \mathbf{X}_1^\top\end{pmatrix}\mathbf{U}^\top \\
&= \begin{pmatrix}\mathbf{U}\mathbf{Z}_1 & \boldsymbol{\Lambda}_1^{\frac{1}{2}}\mathbf{U}\mathbf{Y}_1\end{pmatrix}\begin{pmatrix}\mathbf{O} & \frac{\lambda_i\mathbf{C}(\lambda_i)}{n} \\ \frac{\lambda_i\mathbf{D}(\lambda_i)}{T} & \mathbf{O}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_1^\top\mathbf{U}^\top \\ \mathbf{Y}_1^\top\mathbf{U}^\top\boldsymbol{\Lambda}_1^{\frac{1}{2}}\end{pmatrix}.
\end{aligned}$$

Then, for all $1 \le j_1, j_2 \le q$,

$$\begin{aligned}
&\mathbf{e}_{j_1}^\top\mathbf{U}\left\{\lambda_i \frac{\mathbf{Z}_1\mathbf{C}(\lambda_i)\mathbf{X}_1^\top}{n} + \lambda_i \frac{\mathbf{X}_1\mathbf{D}(\lambda_i)\mathbf{Z}_1^\top}{T}\right\}\mathbf{U}^\top\mathbf{e}_{j_2} = \mathbf{e}_{j_1}^\top\begin{pmatrix}\mathbf{U}\mathbf{Z}_1 & \Lambda_1^{\frac{1}{2}}\mathbf{U}\mathbf{Y}_1\end{pmatrix}\begin{pmatrix}\mathbf{O} & \frac{\lambda_i\mathbf{C}(\lambda_i)}{n} \\ \frac{\lambda_i\mathbf{D}(\lambda_i)}{T} & \mathbf{O}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_1^\top\mathbf{U}^\top \\ \mathbf{Y}_1^\top\mathbf{U}^\top\Lambda_1^{\frac{1}{2}}\end{pmatrix}\mathbf{e}_{j_2} \\
&= \mathbf{e}_{j_1}^\top\begin{pmatrix}\mathbf{U}\mathbf{Z}_1 & \lambda_{j_1}^{\frac{1}{2}}\mathbf{U}\mathbf{Y}_1\end{pmatrix}\begin{pmatrix}\mathbf{O} & \frac{\lambda_i\mathbf{C}(\lambda_i)}{n} \\ \frac{\lambda_i\mathbf{D}(\lambda_i)}{T} & \mathbf{O}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_1^\top\mathbf{U}^\top \\ \lambda_{j_2}^{\frac{1}{2}}\mathbf{Y}_1^\top\mathbf{U}^\top\end{pmatrix}\mathbf{e}_{j_2} = \frac{\lambda_{j_1}^{\frac{1}{2}} + \lambda_{j_2}^{\frac{1}{2}}}{2}\mathbf{e}_{j_1}^\top\begin{pmatrix}\mathbf{U}\mathbf{Z}_1 & \mathbf{U}\mathbf{Y}_1\end{pmatrix}\begin{pmatrix}\mathbf{O} & \frac{\lambda_i\mathbf{C}(\lambda_i)}{n} \\ \frac{\lambda_i\mathbf{D}(\lambda_i)}{T} & \mathbf{O}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_1^\top\mathbf{U}^\top \\ \mathbf{Y}_1^\top\mathbf{U}^\top\end{pmatrix}\mathbf{e}_{j_2} \\
&\quad + \frac{\lambda_{j_1}^{\frac{1}{2}} - \lambda_{j_2}^{\frac{1}{2}}}{2\mathrm{i}}\mathbf{e}_{j_1}^\top\begin{pmatrix}\mathbf{U}\mathbf{Z}_1 & \mathbf{U}\mathbf{Y}_1\end{pmatrix}\begin{pmatrix}\mathbf{O} & \frac{\lambda_i\mathbf{C}(\lambda_i)}{n}\mathrm{i} \\ -\frac{\lambda_i\mathbf{D}(\lambda_i)}{T}\mathrm{i} & \mathbf{O}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_1^\top\mathbf{U}^\top \\ \mathbf{Y}_1^\top\mathbf{U}^\top\end{pmatrix}\mathbf{e}_{j_2} \\
&= \frac{\lambda_{j_1}^{\frac{1}{2}} + \lambda_{j_2}^{\frac{1}{2}}}{2}O_p\left(n^{-\frac{1}{2}}\right) + \frac{\lambda_{j_1}^{\frac{1}{2}} - \lambda_{j_2}^{\frac{1}{2}}}{2\mathrm{i}}O_p\left(n^{-\frac{1}{2}}\right) = \left(\lambda_{j_1}^{\frac{1}{2}} + \lambda_{j_2}^{\frac{1}{2}}\right)O_p\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where $\mathrm{i} := \sqrt{-1}$ is the imaginary unit and the penultimate equality is implied by Theorem 7.1 in [3]. Due to the fact that

$$\begin{aligned}
&\hat{\lambda}_i\mathbf{C}(\hat{\lambda}_i) - \lambda_i\mathbf{C}(\lambda_i) = \mathbf{Z}_2^\top\left\{\mathbf{M}^{-1}(\hat{\lambda}_i) - \mathbf{M}^{-1}(\lambda_i)\right\}\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2 \\
&= \mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\left\{\mathbf{M}(\lambda_i) - \mathbf{M}(\hat{\lambda}_i)\right\}\mathbf{M}^{-1}(\lambda_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2 = \left(\hat{\lambda}_i^{-1} - \lambda_i^{-1}\right)\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{F}_0\mathbf{M}^{-1}(\lambda_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2, \\
&\hat{\lambda}_i\mathbf{D}(\hat{\lambda}_i) - \lambda_i\mathbf{D}(\lambda_i) = \mathbf{X}_2^\top\left\{\mathbf{M}^{-1}(\hat{\lambda}_i) - \mathbf{M}^{-1}(\lambda_i)\right\}\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2 \\
&= \mathbf{X}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\left\{\mathbf{M}(\lambda_i) - \mathbf{M}(\hat{\lambda}_i)\right\}\mathbf{M}^{-1}(\lambda_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2 = \left(\hat{\lambda}_i^{-1} - \lambda_i^{-1}\right)\mathbf{X}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{F}_0\mathbf{M}^{-1}(\lambda_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2,
\end{aligned}$$

we can get

$$\mathbf{e}_{j_1}^\top\mathbf{U}\mathbf{Z}_1\left\{\hat{\lambda}_i\mathbf{C}(\hat{\lambda}_i) - \lambda_i\mathbf{C}(\lambda_i)\right\}\frac{1}{n}\mathbf{Y}_1^\top\mathbf{U}^\top\Lambda_1^{\frac{1}{2}}\mathbf{e}_{j_2} = \lambda_{j_2}^{\frac{1}{2}}O_{\text{a.s.}}\left(\lambda_i^{-1}\right)$$

and

$$\mathbf{e}_{j_1}^\top\Lambda_1^{\frac{1}{2}}\mathbf{U}\mathbf{Y}_1\left\{\hat{\lambda}_i\mathbf{D}(\hat{\lambda}_i) - \lambda_i\mathbf{D}(\lambda_i)\right\}\frac{1}{T}\mathbf{Z}_1^\top\mathbf{U}^\top\mathbf{e}_{j_2} = \lambda_{j_1}^{\frac{1}{2}}O_{\text{a.s.}}\left(\lambda_i^{-1}\right)$$

for any $1 \le j_1, j_2 \le q$. By using the similar arguments for proving (29) and (30), it holds that

$$\mathbf{e}_{j_1}^\top\mathbf{U}(\boldsymbol{\Xi}_{\mathbf{C}} + \boldsymbol{\Xi}_{\mathbf{D}})\mathbf{U}^\top\mathbf{e}_{j_2} = \left(\lambda_{j_1}^{\frac{1}{2}} + \lambda_{j_2}^{\frac{1}{2}}\right)O_p\left(n^{-\frac{1}{2}}\right) + \left(\lambda_{j_1}^{\frac{1}{2}} + \lambda_{j_2}^{\frac{1}{2}}\right)O_{\text{a.s.}}\left(\lambda_i^{-1}\right) \tag{33}$$

for any $1 \le j_1, j_2 \le q$.

Combining (30)–(33) and the determinant in (21), we can compute the limit of $\hat{\lambda}_i/\lambda_i$ for each $1 \le i \le q$. We use a new notation to denote the matrix in the determinant in (21). Define

$$\boldsymbol{\Xi} := \mathbf{U}\left(\boldsymbol{\Xi}_{\mathbf{A}} - \boldsymbol{\Xi}_{\mathbf{B}} + \boldsymbol{\Xi}_{\mathbf{C}} + \boldsymbol{\Xi}_{\mathbf{D}}\right)\mathbf{U}^{\top}, \quad \widetilde{\boldsymbol{\Xi}} := \operatorname{diag}\left(\xi_{11}, \ldots, \xi_{qq}\right),$$

where $\xi_{jj} = \hat{\lambda}_i \{1 - (p-q)/n\} - \lambda_j$. Then by (30)–(33), we have that

$$\begin{aligned}
\mathbf{e}_{j_1}^{\top}\left(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}}\right)\mathbf{e}_{j_2} =& \lambda_i \mathrm{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right) + \mathrm{O}_{\mathrm{p}}(1) + \lambda_{j_1}^{\frac{1}{2}}\lambda_{j_2}^{\frac{1}{2}}\mathrm{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right) + \lambda_{j_1}^{\frac{1}{2}}\lambda_{j_2}^{\frac{1}{2}}\mathrm{O}_{\mathrm{a.s.}}(\lambda_i^{-1}) \\
&+ \left(\lambda_{j_1}^{\frac{1}{2}} + \lambda_{j_2}^{\frac{1}{2}}\right)\mathrm{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right) + \left(\lambda_{j_1}^{\frac{1}{2}} + \lambda_{j_2}^{\frac{1}{2}}\right)\mathrm{O}_{\mathrm{a.s.}}\left(\lambda_i^{-1}\right) \\
=& \left(\lambda_i + \lambda_{j_1}^{\frac{1}{2}}\lambda_{j_2}^{\frac{1}{2}}\right)\left\{\mathrm{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right) + \mathrm{O}_{\mathrm{a.s.}}\left(\lambda_i^{-1}\right)\right\}
\end{aligned}$$

for any $1 \le j_1, j_2 \le q$, which follows that

$$\mathbf{e}_{j_1}^{\top}\lambda_i^{-1}\left(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}}\right)\mathbf{e}_{j_2} = \left(1 + \lambda_i^{-1}\lambda_{j_1}^{\frac{1}{2}}\lambda_{j_2}^{\frac{1}{2}}\right)\left\{\mathrm{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right) + \mathrm{O}_{\mathrm{a.s.}}\left(\lambda_i^{-1}\right)\right\}. \tag{34}$$

According to (23) and (25) for $\boldsymbol{\Xi}_{\mathbf{A}}$ (similar results also hold for $\boldsymbol{\Xi}_{\mathbf{B}}$, $\boldsymbol{\Xi}_{\mathbf{C}}$ and $\boldsymbol{\Xi}_{\mathbf{D}}$), it can be checked that the variance of the term in (34) has the order

$$\left(1 + \lambda_i^{-1}\lambda_{j_1}^{\frac{1}{2}}\lambda_{j_2}^{\frac{1}{2}}\right)^2\left(n^{-\frac{1}{2}} + \lambda_i^{-1}\right)^2.$$

By Chebyshev's inequality, we have that, for any $\epsilon > 0$,

$$\begin{aligned}
&\Pr\left\{\max_{1\le j_1, j_2\le q}\left|\mathbf{e}_{j_1}^{\top}\lambda_i^{-1}\left(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}}\right)\mathbf{e}_{j_2}\right| \ge \epsilon\left(n^{-\frac{1}{2}} + \lambda_i^{-1}\right)\right\} \le \sum_{1\le j_1, j_2\le q}\Pr\left\{\left|\mathbf{e}_{j_1}^{\top}\lambda_i^{-1}\left(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}}\right)\mathbf{e}_{j_2}\right| \ge \epsilon\left(n^{-\frac{1}{2}} + \lambda_i^{-1}\right)\right\} \\
&\le \sum_{1\le j_1, j_2\le q}\frac{\mathrm{E}\left\{\mathbf{e}_{j_1}^{\top}\lambda_i^{-1}\left(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}}\right)\mathbf{e}_{j_2}\right\}^2}{\epsilon^2\left(n^{-\frac{1}{2}} + \lambda_i^{-1}\right)^2} = \sum_{1\le j_1, j_2\le q}\left(1 + \lambda_i^{-1}\lambda_{j_1}^{\frac{1}{2}}\lambda_{j_2}^{\frac{1}{2}}\right)^2\mathrm{O}\left(\epsilon^{-2}\right) \\
&= \left(q + \lambda_i^{-1}\sum_{j=1}^{q}\lambda_j\right)^2\mathrm{O}\left(\epsilon^{-2}\right) = \kappa_1^2\mathrm{O}\left(\epsilon^{-2}\right),
\end{aligned}$$

which means

$$\|\lambda_i^{-1}(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}})\|_{\infty} = \max_{1\le j_1, j_2\le q}\left|\mathbf{e}_{j_1}^{\top}\lambda_i^{-1}\left(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}}\right)\mathbf{e}_{j_2}\right| = \kappa_1\mathrm{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}} + \lambda_i^{-1}\right)$$

and then

$$|||\lambda_i^{-1}(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}})|||_{\infty} \le q\|\lambda_i^{-1}(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}})\|_{\infty} = \kappa_1 q\mathrm{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}} + \lambda_i^{-1}\right).$$

Note that the determinant equation $\det(\widetilde{\boldsymbol{\Xi}}) = 0$ is equivalent to $\det\left(\lambda_i^{-1}\widetilde{\boldsymbol{\Xi}}\right) = 0$, that is,

$$\det\left\{\frac{\hat{\lambda}_i}{\lambda_i}\left(1 - \frac{p-q}{n}\right)\mathbf{I}_q - \lambda_i^{-1}\boldsymbol{\Lambda}_1\right\} = 0.$$

At the same time, the equation $\det(\boldsymbol{\Xi}) = 0$ is equivalent to $\det\left(\lambda_i^{-1}\boldsymbol{\Xi}\right) = 0$, that is,

$$\det\left\{\frac{\hat{\lambda}_i}{\lambda_i}\left(1 - \frac{p-q}{n}\right)\mathbf{I}_q - \lambda_i^{-1}\boldsymbol{\Lambda}_1 + \lambda_i^{-1}\left(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}}\right)\right\} = 0.$$

By eigenvalue perturbation theorems (see Theorem 6.3.2 in Chapter 6, [9]), we have

$$\left|\frac{\hat{\lambda}_i}{\lambda_i}\left(1 - \frac{p-q}{n}\right) - 1\right| \le |||\lambda_i^{-1}(\boldsymbol{\Xi} - \widetilde{\boldsymbol{\Xi}})|||_{\infty} = \kappa_1 q\mathrm{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}} + \lambda_i^{-1}\right),$$

that is

$$\frac{\hat{\lambda}_i}{\lambda_i} = \frac{1}{1-y} + \mathrm{O}\left(y_p - y\right) + \kappa_1 q\mathrm{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}} + \lambda_i^{-1}\right). \tag{35}$$

Instead, we can compare the determinant equations $\det\left(\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\widetilde{\boldsymbol{\Xi}}\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\right)=0$ and $\det\left(\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\boldsymbol{\Xi}\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\right)=0$, and then repeat all the derivations above to achieve an upper bound of $\|\boldsymbol{\Lambda}_1^{-1/2}(\boldsymbol{\Xi}-\widetilde{\boldsymbol{\Xi}})\boldsymbol{\Lambda}_1^{-1/2}\|_\infty$. In this case, we can get

$$\frac{\hat{\lambda}_i}{\lambda_i}=\frac{1}{1-y}+\mathrm{O}\left(y_p-y\right)+\kappa_2 q\mathrm{O}_\mathrm{p}\left(n^{-\frac{1}{2}}+\lambda_i^{-1}\right). \tag{36}$$

Thus, (35) and (36) lead to

$$\frac{\hat{\lambda}_i}{\lambda_i}=\frac{1}{1-y}+\mathrm{O}\left(y_p-y\right)+\kappa q\mathrm{O}_\mathrm{p}\left(n^{-\frac{1}{2}}+\lambda_i^{-1}\right),$$

where $\kappa q(n^{-1/2}+\lambda_i^{-1})=\mathrm{o}(1)$ under Assumption 2. The proof is finished. □

Proof of Theorem 2.2. We begin with the equation on $\hat{\lambda}_i$ in (20). Recall that we have expressed (20) as

$$\det\left(\boldsymbol{\Xi}_\mathbf{A}-\boldsymbol{\Xi}_\mathbf{B}+\boldsymbol{\Xi}_\mathbf{C}+\boldsymbol{\Xi}_\mathbf{D}\right)=0. \tag{37}$$

For the first term $\boldsymbol{\Xi}_\mathbf{A}$, we can write it as

$$\begin{aligned}\boldsymbol{\Xi}_\mathbf{A}=&\frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\left[\left\{\mathbf{I}_n-\mathbf{A}(\hat{\lambda}_i)\right\}-\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\right]\mathbf{Z}_1^\top+\frac{\hat{\lambda}_i}{n}\left(\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top-\mathrm{E}\left[\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right]\right)\\&+\frac{\hat{\lambda}_i}{n}\mathrm{E}\left[\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right].\end{aligned}$$

Using the fact

$$\left\{\mathbf{I}_n-\mathbf{A}(\hat{\lambda}_i)\right\}-\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}=-\delta_i\mathbf{A}(\theta_i)+\delta_i\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2,$$

$\hat{\lambda}_i=\theta_i(1+\delta_i)$ and (13), we have

$$\begin{aligned}\boldsymbol{\Xi}_\mathbf{A}=&\,\theta_i\delta_i(1+\delta_i)\frac{1}{n}\left(\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top-\mathrm{E}\left[\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right]\right)\\&+\theta_i\delta_i(1+\delta_i)\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top\\&-\theta_i\delta_i(1+\delta_i)\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_1^\top+\theta_i\delta_i(1+\delta_i)\frac{1}{n}\mathrm{E}\left[\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right]\\&+\theta_i(1+\delta_i)\frac{1}{n}\left(\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top-\mathrm{E}\left[\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right]\right)\\&+\theta_i(1+\delta_i)\frac{1}{n}\mathrm{E}\left[\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right]\\=&\,\theta_i(1+\delta_i)^2\frac{1}{n}\left(\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top-\mathrm{E}\left[\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right]\right)\\&+\theta_i\delta_i(1+\delta_i)\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top\\&-\theta_i\delta_i(1+\delta_i)\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_1^\top+\theta_i(1+\delta_i)^2\frac{1}{n}\mathrm{E}\left[\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right]\\=:&\,\theta_i(1+\delta_i)^2\boldsymbol{\Xi}_{\mathbf{A}1}+\theta_i\delta_i(1+\delta_i)\boldsymbol{\Xi}_{\mathbf{A}2}-\theta_i\delta_i(1+\delta_i)\boldsymbol{\Xi}_{\mathbf{A}3}+\theta_i(1+\delta_i)^2\boldsymbol{\Xi}_{\mathbf{A}4}.\end{aligned} \tag{38}$$

For the second term $\boldsymbol{\Xi}_\mathbf{B}$, we can similarly write it as

$$\begin{aligned}\boldsymbol{\Xi}_\mathbf{B}=&\frac{1}{T}\mathbf{X}_1\left\{\mathbf{B}(\hat{\lambda}_i)-\mathbf{B}(\theta_i)\right\}\mathbf{X}_1^\top+\frac{1}{T}\left(\mathbf{X}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{X}_1^\top-\mathrm{E}\left[\mathbf{X}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{X}_1^\top\right]\right)+\frac{1}{T}\mathrm{E}\left[\mathbf{X}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{X}_1^\top\right]\\=&\frac{1}{T}\left(\mathbf{X}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{X}_1^\top-\mathrm{E}\left[\mathbf{X}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{X}_1^\top\right]\right)-\frac{\delta_i}{\hat{\lambda}_i}\frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top\\&+\frac{1}{T}\mathrm{E}\left[\mathbf{X}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{X}_1^\top\right]=:\boldsymbol{\Xi}_{\mathbf{B}1}-\frac{\delta_i}{\hat{\lambda}_i}\boldsymbol{\Xi}_{\mathbf{B}2}+\boldsymbol{\Xi}_{\mathbf{B}3},\end{aligned} \tag{39}$$

where the second equality above uses the fact

$$\mathbf{B}(\hat{\lambda}_i)-\mathbf{B}(\theta_i)=-\frac{\delta_i}{\hat{\lambda}_i}\mathbf{X}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2.$$

For the term $\Xi_{\mathbf{C}}$, we have

$$\Xi_{\mathbf{C}}=\frac{\hat{\lambda}_i}{n}\mathbf{Z}_1\left\{\mathbf{C}(\hat{\lambda}_i)-\mathbf{C}(\theta_i)\right\}\mathbf{X}_1^\top+\frac{\hat{\lambda}_i-\theta_i}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top+\frac{\theta_i}{n}\left[\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top-\mathrm{E}\left\{\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top\right\}\right].$$

Using the fact

$$\mathbf{C}(\hat{\lambda}_i)-\mathbf{C}(\theta_i)=-\frac{\delta_i}{\hat{\lambda}_i}\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2,$$

we have the decomposition

$$\begin{aligned}\Xi_{\mathbf{C}}&=\theta_i\frac{1}{n}\left[\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top-\mathrm{E}\left\{\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top\right\}\right]-\delta_i\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top+\theta_i\delta_i\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top\\&=:\theta_i\Xi_{\mathbf{C}1}-\delta_i\Xi_{\mathbf{C}2}+\theta_i\delta_i\Xi_{\mathbf{C}3}.\end{aligned}\tag{40}$$

Similarly, we can write the last term $\Xi_{\mathbf{D}}$ as

$$\begin{aligned}\Xi_{\mathbf{D}}&=\theta_i\frac{1}{T}\left[\mathbf{X}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top-\mathrm{E}\left\{\mathbf{X}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top\right\}\right]-\delta_i\frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top\mathbf{M}^{-1}(\hat{\lambda}_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top+\theta_i\delta_i\frac{1}{T}\mathbf{X}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top\\&=:\theta_i\Xi_{\mathbf{D}1}-\delta_i\Xi_{\mathbf{D}2}+\theta_i\delta_i\Xi_{\mathbf{D}3}.\end{aligned}\tag{41}$$

Putting (38)–(41) into (37), we have

$$\det(\theta_i\mathbf{\Theta}_{1n}+\theta_i\delta_i\mathbf{\Theta}_{2n}+\theta_i\mathbf{\Theta}_{3n})=0,\tag{42}$$

where

$$\mathbf{\Theta}_{1n}:=(1+\delta_i)^2\Xi_{\mathbf{A}1}-\theta_i^{-1}\Xi_{\mathbf{B}1}+\Xi_{\mathbf{C}1}+\Xi_{\mathbf{D}1},\tag{43}$$

$$\mathbf{\Theta}_{2n}:=(1+\delta_i)\Xi_{\mathbf{A}2}-(1+\delta_i)\Xi_{\mathbf{A}3}+\frac{1}{\theta_i\hat{\lambda}_i}\Xi_{\mathbf{B}2}-\theta_i^{-1}\Xi_{\mathbf{C}2}+\Xi_{\mathbf{C}3}-\theta_i^{-1}\Xi_{\mathbf{D}2}+\Xi_{\mathbf{D}3},\tag{44}$$

$$\mathbf{\Theta}_{3n}:=(1+\delta_i)^2\Xi_{\mathbf{A}4}-\theta_i^{-1}\Xi_{\mathbf{B}3}.\tag{45}$$

Multiplying both sides of the matrix in (42) by $\theta_i^{-1/2}\mathbf{U}$ from the left-hand side and $\theta_i^{-1/2}\mathbf{U}^\top$ from the right-hand side, we get

$$\det\left\{\mathbf{U}(\mathbf{\Theta}_{1n}+\delta_i\mathbf{\Theta}_{2n}+\mathbf{\Theta}_{3n})\mathbf{U}^\top\right\}=0.\tag{46}$$

Introduce the following lemmas for $\mathbf{\Theta}_{1n}$ and $\mathbf{\Theta}_{2n}$.

Lemma 3.2. *For any fixed $1\le i\le q$, denote $\mathbf{G}_{ni}=\sqrt{p}\mathbf{U}\mathbf{\Theta}_{1n}\mathbf{U}^\top$ with $\mathbf{U}$ defined in (6). Under the assumptions of Theorem 2.2,*

$$\mathbf{e}_i^\top\mathbf{G}_{ni}\mathbf{e}_i\xrightarrow{d}\mathcal{N}(0,\tilde{\sigma}_i^2),\tag{47}$$

where $\mathbf{e}_i$ is the q-dimensional vector whose ith element is 1 and otherwise 0, $\tilde{\sigma}_i^2=(y+c)(1-y)^2\nu_i-y(1-y)(1-3y)-c(1-y)^2$, and $\nu_i=\mathbb{E}|\mathbf{u}_i^\top\mathbf{Z}_1\mathbf{e}_1|^4$ for $1\le i\le q$.

Lemma 3.3. *Under the assumptions of Theorem 2.2,*

$$\|\mathbf{U}\mathbf{\Theta}_{1n}\mathbf{U}^\top\|_\infty=\mathrm{O_p}\left(\frac{q}{\sqrt{n}}+\frac{\sum_j\lambda_j}{\sqrt{n}\lambda_i}\right).\tag{48}$$

Lemma 3.4. *Under the assumptions of Theorem 2.2,*

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top\mathbf{U}\mathbf{\Theta}_{2n}\mathbf{U}^\top\mathbf{e}_j-(y-1)\right|=\mathrm{O_p}\left(\frac{\sqrt{q}\delta_i}{\lambda_i}+\frac{\sqrt{q}}{\sqrt{n}}+\frac{\sqrt{\sum_j\lambda_j^2}}{\lambda_i^2}+\frac{\sqrt{\sum_j\lambda_j}}{\lambda_i}\right),\tag{49}$$

$$\max_{1\le j_1\ne j_2\le q}\left|\mathbf{e}_{j_1}^\top\mathbf{U}\mathbf{\Theta}_{2n}\mathbf{U}^\top\mathbf{e}_{j_2}\right|=\mathrm{O_p}\left(\frac{q\delta_i}{\lambda_i}+\frac{q}{\sqrt{n}}+\frac{\sum_j\lambda_j}{\lambda_i^2}+\frac{\sqrt{q\sum_j\lambda_j}}{\lambda_i}\right).\tag{50}$$

For the term $\mathbf{U}\mathbf{\Theta}_{3n}\mathbf{U}^\top$ in (46), by considering its (j_1,j_2) entry for all $1\le j_1,j_2\le q$, we can get that

$$(1+\delta_i)^2\mathbf{U}\Xi_{\mathbf{A}4}\mathbf{U}^\top=(1+\delta_i)^2\left[1-\frac{p-q}{n}\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}\right]\mathbf{I}_q,\quad\mathbf{U}\Xi_{\mathbf{B}3}\mathbf{U}^\top=\left(1+\frac{p-q}{T}\left[-1+\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}\right]\right)\mathbf{\Lambda}_1.\tag{51}$$

By the definition of θ_i in (12), we know

$$1-\frac{p-q}{n}\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}=\frac{\lambda_i}{\theta_i}\left(1+\frac{p-q}{T}\left[-1+\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}\right]\right),$$

which, together with the results in Lemma 3.1 and Theorem 2.1, yields that

$$\begin{aligned}&(1+\delta_i)^2\left[1-\frac{p-q}{n}\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}\right]-\frac{\lambda_i}{\theta_i}\left(1+\frac{p-q}{T}\left[-1+\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}\right]\right)\\&=2\delta_i\left[1-\frac{p-q}{n}\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}\right]+\delta_i^2\left[1-\frac{p-q}{n}\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}\right]=2\delta_i\left\{1-\frac{p-q}{n}+\mathrm{o}(1)\right\}.\end{aligned}\tag{52}$$

Combining (51)–(52) and the definition of $\mathbf{\Theta}_{3n}$ in (45), we can get that, for $1\le j\le q$,

$$\mathbf{e}_j^\top\mathbf{U}\mathbf{\Theta}_{3n}\mathbf{U}^\top\mathbf{e}_j^\top=\left\{(1+\delta_i)^2-\frac{\lambda_j}{\lambda_i}\right\}\left[1-\frac{p-q}{n}\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}\right],$$

which converges to zero if and only if $\lambda_j=\lambda_i$ because $(1+\delta_i)^2-\lambda_j/\lambda_i>C>0$ for some constant C if $\lambda_j\neq\lambda_i$ under Assumption 4. When $\lambda_j=\lambda_i$, we have

$$\mathbf{e}_j^\top\mathbf{U}\mathbf{\Theta}_{3n}\mathbf{U}^\top\mathbf{e}_j^\top=2\delta_i\left\{1-\frac{p-q}{n}+\mathrm{o}(1)\right\}.\tag{53}$$

Note that all off-diagonal entries of the matrix $\mathbf{U}\mathbf{\Theta}_{3n}\mathbf{U}^\top$ are zero, i.e.,

$$\mathbf{e}_{j_1}^\top\mathbf{U}\mathbf{\Theta}_{3n}\mathbf{U}^\top\mathbf{e}_{j_2}=0,\forall 1\le j_1\neq j_2\le q.\tag{54}$$

Inserting (47), (48), (49), (50), (53) and (54) into (46), we can solve (46) and get the limiting distribution of $\delta_i(1\le i\le q)$ immediately. Since the diagonal elements of $\mathbf{U}\mathbf{\Theta}_{3n}\mathbf{U}^\top$ are at least of constant order, when $\mathbf{e}_j^\top\mathbf{U}\mathbf{\Theta}_{3n}\mathbf{U}^\top\mathbf{e}_j^\top$ goes to infinity for some j's, we can divide these rows by $\mathbf{e}_j^\top\mathbf{U}\mathbf{\Theta}_{3n}\mathbf{U}^\top\mathbf{e}_j^\top$. In this way, we can get

$$\det\begin{pmatrix}\mathrm{O_p}(1)&\cdots&\mathrm{O_p}(\psi)&\cdots&\mathrm{O_p}(\psi)\\\vdots&\ddots&\vdots&\ddots&\vdots\\\mathrm{O_p}(\psi)&\cdots&\hat{S}_i+(1-y+\mathrm{o_p}(1))\delta_i&\cdots&\mathrm{O_p}(\psi)\\\vdots&\ddots&\vdots&\ddots&\vdots\\\mathrm{O_p}(\psi)&\cdots&\mathrm{O_p}(\psi)&\cdots&\mathrm{O_p}(1)\end{pmatrix}=0,$$

where $\sqrt{p}\hat{S}_i\xrightarrow{d}\mathcal{N}(0,\tilde{\sigma}_i^2)$ and

$$\psi=\frac{q}{\sqrt{n}}+\frac{\sum_j\lambda_j}{\sqrt{n}\lambda_i}+\frac{q\delta_i^2}{\lambda_i}+\frac{\delta_i\sum_j\lambda_j}{\lambda_i^2}+\frac{\delta_i\sqrt{q\sum_j\lambda_j}}{\lambda_i}.$$

By Leibniz formula for determinants, we can get that $\hat{S}_i+\left\{1-y+\mathrm{o_p}(1)\right\}\delta_i+q\mathrm{O_p}\left(\psi^2\right)=0$, that is

$$\hat{S}_i+\left\{1-y+\mathrm{o_p}(1)\right\}\delta_i+\mathrm{O_p}\left(\frac{q^3}{n}+\frac{q(\sum_j\lambda_j)^2}{n\lambda_i^2}+\frac{q^3\delta_i^4}{\lambda_i^2}+\frac{q\delta_i^2(\sum_j\lambda_j)^2}{\lambda_i^4}+\frac{q^2\delta_i^2\sum_j\lambda_j}{\lambda_i^2}\right)=0.$$

Under Assumptions 1 and 2(a), we have $q=\mathrm{o}(n^{\frac{1}{6}})$ and $\lambda_i^{-1}\sum_j\lambda_j=\mathrm{o}(q^{-\frac{1}{2}}n^{\frac{1}{4}})$, then it follows that

$$\frac{q^3}{n}=\mathrm{o}(n^{-\frac{1}{2}}),\quad\frac{q^2\sum_j\lambda_j}{n\lambda_i^2}=\mathrm{o}(n^{-\frac{1}{2}}),\quad\frac{q^3\delta_i^4}{\lambda_i^2}=\mathrm{o_p}(\delta_i^2n^{\frac{1}{2}}),\quad\frac{q\delta_i^2(\sum_j\lambda_j)^2}{\lambda_i^4}=\mathrm{o_p}(\delta_i^2n^{\frac{1}{2}}),\quad\frac{q^2\delta_i^2\sum_j\lambda_j}{\lambda_i^2}=\mathrm{o_p}(\delta_i^2n^{\frac{1}{2}}).$$

It leads to

$$\hat{S}_i+\left\{1-y+\mathrm{o_p}(1)\right\}\delta_i+\mathrm{o_p}(\delta_i^2n^{\frac{1}{2}})+\mathrm{o}(n^{-\frac{1}{2}})=0.$$

By multiplying $\sqrt{p}$ on both sides, we further obtain that

$$\sqrt{p}\hat{S}_i+\left\{1-y+\mathrm{o_p}(1)\right\}\sqrt{p}\delta_i+\mathrm{o_p}(1)\,p\delta_i^2+\mathrm{o}(1)=0.$$

Recalling that $\sqrt{p}\hat{S}_i\xrightarrow{d}\mathcal{N}(0,\tilde{\sigma}_i^2)$, we can achieve $\sqrt{p}\delta_i\xrightarrow{d}\mathcal{N}(0,\sigma_i^2)$, where

$$\sigma_i^2=\frac{\tilde{\sigma}_i^2}{(1-y)^2}=(y+c)\nu_i-c-\frac{y(1-3y)}{1-y}.$$

Instead, we can consider the determinant

$$\det\left\{\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\mathbf{U}(\theta_i\boldsymbol{\Theta}_{1n}+\theta_i\delta_i\boldsymbol{\Theta}_{2n}+\theta_i\boldsymbol{\Theta}_{3n})\mathbf{U}^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\right\}=0, \tag{55}$$

where $\tilde{\boldsymbol{\Lambda}}=\mathrm{diag}(\theta_1,\ldots,\theta_q)\in\mathbb{R}^{q\times q}$. Repeating all the derivations above, we can get

$$\|\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\mathbf{U}\theta_i\boldsymbol{\Theta}_{1n}\mathbf{U}^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\|_\infty=\mathrm{O}_\mathrm{p}\left(\frac{q}{\sqrt{n}}+\frac{\lambda_i\sum_j\lambda_j^{-1}}{\sqrt{n}}\right), \tag{56}$$

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\theta_i\mathbf{U}\boldsymbol{\Theta}_{2n}\mathbf{U}^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\mathbf{e}_j-(y-1)\frac{\theta_i}{\theta_j}\right|=\mathrm{O}_\mathrm{p}\left(\delta_i\sqrt{\sum_j\lambda_j^{-2}}+\frac{\lambda_i\sqrt{\sum_j\lambda_j^{-2}}}{\sqrt{n}}+\frac{\sqrt{q}}{\lambda_i}+\sqrt{\sum_j\lambda_j^{-1}}\right), \tag{57}$$

$$\max_{1\le j_1\neq j_2\le q}\left|\mathbf{e}_{j_1}^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\theta_i\mathbf{U}\boldsymbol{\Theta}_{2n}\mathbf{U}^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\mathbf{e}_{j_2}\right|=\mathrm{O}_\mathrm{p}\left(\delta_i\sum_j\lambda_j^{-1}+\frac{\lambda_i\sum_j\lambda_j^{-1}}{\sqrt{n}}+\frac{q}{\lambda_i}+\sqrt{q\sum_j\lambda_j^{-1}}\right), \tag{58}$$

$$\mathbf{e}_j^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\theta_i\mathbf{U}\boldsymbol{\Theta}_{3n}\mathbf{U}^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\mathbf{e}_j^\top=\left\{(1+\delta_i)^2-\frac{\lambda_j}{\lambda_i}\right\}\left[1-\frac{p-q}{n}\mathrm{E}\left\{\tilde{m}_{\theta_i}(1)\right\}\right], \tag{59}$$

$$\mathbf{e}_{j_1}^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\theta_i\mathbf{U}\boldsymbol{\Theta}_{3n}\mathbf{U}^\top\tilde{\boldsymbol{\Lambda}}^{-\frac{1}{2}}\mathbf{e}_{j_2}=0,\ \forall 1\le j_1\neq j_2\le q. \tag{60}$$

Inserting (56)–(60) into (55), we can similarly prove $\sqrt{p}\delta_i\xrightarrow{d}\mathcal{N}(0,\sigma_i^2)$ under Assumption 2(b). Thus the proof is completed. □

Proof of Theorem 2.3. The proof of Theorem 2.3 is similar to that of Theorem 2.2, the only difference is that we take the $J_i\times J_i$ block as a typical object to analyze, some useful lemmas can also be obtained from Lemmas 3.2–3.4. Similar arguments for deriving Theorem 4.1 in [17] can be used. Thus, we omit the details. □

4. Proofs of technical lemmas

Proof of Lemma 3.1. By the definition of $\widetilde{m}_\theta(z)$ in (8),

$$\widetilde{m}_\theta(1)=\frac{1}{p-q}\mathrm{tr}\left(\mathbf{I}_{p-q}-\frac{\mathbf{F}_0}{\theta}\right)^{-1}=1+\frac{1}{p-q}\mathrm{tr}\left\{\frac{\mathbf{F}_0}{\theta}\left(\mathbf{I}_{p-q}-\frac{\mathbf{F}_0}{\theta}\right)^{-1}\right\},$$

we have

$$\widetilde{m}_\theta(1)-1=\frac{1}{p-q}\mathrm{tr}\left\{\frac{\mathbf{F}_0}{\theta}\left(\mathbf{I}_{p-q}-\frac{\mathbf{F}_0}{\theta}\right)^{-1}\right\}=\theta^{-1}\left(\frac{1}{p-q}\sum_{1\le j\le p-q}\frac{\mu_j}{1-\mu_j/\theta}\right).$$

Since all the eigenvalues of $\mathbf{F}_0$, namely $\mu_1\ge\cdots\ge\mu_{p-q}$, are almost surely bounded, we can get that $\widetilde{m}_\theta(1)-1=\mathrm{O}_{\mathrm{a.s.}}(\theta^{-1})$. □

Proof of Lemma 3.2. From the definition of $\boldsymbol{\Theta}_{1n}$ in (43) and the fact that $\mathbf{Y}_1=\boldsymbol{\Sigma}_1^{-\frac{1}{2}}\mathbf{X}_1=\mathbf{U}^\top\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}\mathbf{X}_1$, we have the decomposition

$$\begin{aligned}\mathbf{e}_i^\top\mathbf{G}_{ni}\mathbf{e}_i=\mathbf{u}_i^\top\Big[&\frac{(1+\delta_i)^2\sqrt{p}}{n}\mathbf{Z}_1\left\{\mathbf{I}_n-\mathbf{A}(\theta_i)\right\}\mathbf{Z}_1^\top-\frac{\lambda_i}{\theta_i}\frac{\sqrt{p}}{T}\mathbf{Y}_1\left\{\mathbf{I}_T+\mathbf{B}(\theta_i)\right\}\mathbf{Y}_1^\top+\frac{\sqrt{p\lambda_i}}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{Y}_1^\top+\frac{\sqrt{p\lambda_i}}{T}\mathbf{Y}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top\Big]\mathbf{u}_i\\&-\mathrm{E}(\cdot),\end{aligned} \tag{61}$$

where $\mathrm{E}[\cdot]$ is the expectation of all the preceding terms after the equal sign.

By Theorem 2.1, δ_i converges in probability to 0, thus we only need to consider the limit of

$$\mathbf{e}_i^\top\widetilde{\mathbf{G}}_{ni}\mathbf{e}_i:=\mathbf{u}_i^\top\Big[\frac{\sqrt{p}}{n}\mathbf{Z}_1\left\{\mathbf{I}_n-\mathbf{A}(\theta_i)\right\}\mathbf{Z}_1^\top-\frac{\lambda_i}{\theta_i}\frac{\sqrt{p}}{T}\mathbf{Y}_1\left\{\mathbf{I}_T+\mathbf{B}(\theta_i)\right\}\mathbf{Y}_1^\top+\frac{\sqrt{p\lambda_i}}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{Y}_1^\top+\frac{\sqrt{p\lambda_i}}{T}\mathbf{Y}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top\Big]\mathbf{u}_i-\mathrm{E}[\cdot].$$

For the first two terms, Theorem 7.2 in [3] implies that, for any $1\le i\le q$,

$$\frac{1}{\sqrt{n}}\left[\mathbf{u}_i^\top\mathbf{Z}_1\left\{\mathbf{I}_n-\mathbf{A}(\theta_i)\right\}\mathbf{Z}_1^\top\mathbf{u}_i-\mathrm{tr}\left\{\mathbf{I}_n-\mathbf{A}(\theta_i)\right\}\right]\xrightarrow{d}\mathcal{N}(0,\tilde{\sigma}_{i\mathbf{A}}^2),$$

$$\frac{1}{\sqrt{T}}\left[\mathbf{u}_i^\top\mathbf{Y}_1\left\{\mathbf{I}_T+\mathbf{B}(\theta_i)\right\}\mathbf{Y}_1^\top\mathbf{u}_i-\mathrm{tr}\left\{\mathbf{I}_T+\mathbf{B}(\theta_i)\right\}\right]\xrightarrow{d}\mathcal{N}(0,\tilde{\sigma}_{i\mathbf{B}}^2),$$

with $\tilde{\sigma}_{iA}^2 = \omega_{\mathbf{I}_n-\mathbf{A}(\theta_i)}(\nu_i - 3) + 2\beta_{\mathbf{I}_n-\mathbf{A}(\theta_i)}$ and $\tilde{\sigma}_{iB}^2 = \omega_{\mathbf{I}_T+\mathbf{B}(\theta_i)}(\nu_i - 3) + 2\beta_{\mathbf{I}_T+\mathbf{B}(\theta_i)}$, where

$$\nu_i = \mathrm{E}|\mathbf{u}_i^\top \mathbf{Z}_1 \mathbf{e}_1|^4 = \mathrm{E}|\mathbf{u}_i^\top \mathbf{Y}_1 \mathbf{e}_1|^4,\ \ \omega_{\mathbf{I}_n-\mathbf{A}(\theta_i)} = \lim_{n\to\infty} \frac{1}{n} \sum_{1\le k\le n} \left[\{\mathbf{I}_n - \mathbf{A}(\theta_i)\}(k,k)\right]^2,\ \ \beta_{\mathbf{I}_n-\mathbf{A}(\theta_i)} = \lim_{n\to\infty} \frac{1}{n} \mathrm{tr}\{\mathbf{I}_n - \mathbf{A}(\theta_i)\}^2,$$

$\omega_{\mathbf{I}_T+\mathbf{B}(\theta_i)}$ and $\beta_{\mathbf{I}_T+\mathbf{B}(\theta_i)}$ are similarly defined. Here the fact that $\mathrm{E}|\mathbf{u}_i^\top \mathbf{Z}_1 \mathbf{e}_1|^4 = \mathrm{E}|\mathbf{u}_i^\top \mathbf{Y}_1 \mathbf{e}_1|^4$ is implied by Assumption 3. Based on the fact that

$$\begin{aligned}\mathrm{E}\left[\mathbf{u}_i^\top \mathbf{Z}_1 \{\mathbf{I}_n - \mathbf{A}(\theta_i)\} \mathbf{Z}_1^\top \mathbf{u}_i\right] &= \mathrm{E}\left(\mathrm{tr}\left[\mathbf{Z}_1^\top \mathbf{u}_i \mathbf{u}_i^\top \mathbf{Z}_1 \{\mathbf{I}_n - \mathbf{A}(\theta_i)\}\right]\right) = \mathrm{tr}\left[\mathrm{E}\left(\mathbf{Z}_1^\top \mathbf{u}_i \mathbf{u}_i^\top \mathbf{Z}_1\right) \mathrm{E}\{\mathbf{I}_n - \mathbf{A}(\theta_i)\}\right]\\ &= \mathrm{E}\left[\mathrm{tr}\{\mathbf{I}_n - \mathbf{A}(\theta_i)\}\right] = n - (p-q)\mathrm{E}\left\{\widetilde{m}_{\theta_i}(1)\right\},\end{aligned}$$

and that $\widetilde{m}_{\theta_i}(1) - \mathrm{E}\left\{\widetilde{m}_{\theta_i}(1)\right\} = \mathrm{O_p}(n^{-1})$, we can get that

$$\frac{1}{\sqrt{n}} \left(\mathrm{E}\left[\mathbf{u}_i^\top \mathbf{Z}_1 \{\mathbf{I}_n - \mathbf{A}(\theta_i)\} \mathbf{Z}_1^\top \mathbf{u}_i^\top\right] - \mathrm{tr}\{\mathbf{I}_n - \mathbf{A}(\theta_i)\}\right) = \mathrm{o_p}(1).$$

Then it follows that

$$\frac{1}{\sqrt{n}} \left(\mathbf{u}_i^\top \mathbf{Z}_1 \{\mathbf{I}_n - \mathbf{A}(\theta_i)\} \mathbf{Z}_1^\top \mathbf{u}_i - \mathrm{E}\left[\mathbf{u}_i^\top \mathbf{Z}_1 \{\mathbf{I}_n - \mathbf{A}(\theta_i)\} \mathbf{Z}_1^\top \mathbf{u}_i^\top\right]\right) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_{i\mathbf{A}}^2),$$

and similarly,

$$\frac{1}{\sqrt{T}} \left(\mathbf{u}_i^\top \mathbf{Y}_1 \{\mathbf{I}_T + \mathbf{B}(\theta_i)\} \mathbf{Y}_1^\top \mathbf{u}_i - \mathrm{E}\left[\mathbf{u}_i^\top \mathbf{Y}_1 \{\mathbf{I}_T + \mathbf{B}(\theta_i)\} \mathbf{Y}_1^\top \mathbf{u}_i\right]\right) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_{i\mathbf{B}}^2).$$

For other two terms, by the same approach in the proof of Theorem 2.1, we have that

$$\mathbf{u}_i^\top \left\{\frac{\sqrt{p\lambda_i}}{n} \mathbf{Z}_1 \mathbf{C}(\theta_i) \mathbf{Y}_1^\top + \frac{\sqrt{p\lambda_i}}{T} \mathbf{Y}_1 \mathbf{D}(\theta_i) \mathbf{Z}_1^\top\right\} \mathbf{u}_i = \mathrm{O_p}\left(\frac{1}{\sqrt{\lambda_i}}\right).$$

By all these arguments above, we can derive that $\mathbf{e}_i^\top \mathbf{G}_{ni} \mathbf{e}_i \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_i^2)$ with $\tilde{\sigma}_i^2 = y\sigma_{i\mathbf{A}}^2 + c(1-y)^2 \sigma_{i\mathbf{B}}^2$.

We compute $\omega_{\mathbf{I}_n-\mathbf{A}(\theta_i)}$, $\beta_{\mathbf{I}_n-\mathbf{A}(\theta_i)}$, $\omega_{\mathbf{I}_T+\mathbf{B}(\theta_i)}$ and $\beta_{\mathbf{I}_T+\mathbf{B}(\theta_i)}$ in the following. By the derivations in the proof of Lemma 6 in [17],

$$\begin{aligned}\{\mathbf{I}_n - \mathbf{A}(\theta_i)\}(k,k) &= 1 - \left\{\mathbf{Z}_2^\top \mathbf{M}(\theta_i)^{-1} \left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1} \frac{1}{n}\mathbf{Z}_2\right\}(k,k) = 1 - \frac{\theta_i}{n}\left\{\mathbf{Z}_2^\top \left(\theta_i \cdot \frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top\right)^{-1} \mathbf{Z}_2\right\}(k,k)\\ &= \frac{1}{1 + \frac{\theta_i}{n}\left\{\eta_k^\top \left(\theta_i \frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top\right)^{-1} \eta_k\right\}},\end{aligned}$$

where η_k is the kth column of $\mathbf{Z}_2$ and $\mathbf{Z}_{2k}$ is defined by removing the kth column of $\mathbf{Z}_2$.

Note that

$$\begin{aligned}&\left(\frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top - \frac{1}{\theta_i T}\mathbf{X}_2\mathbf{X}_2^\top\right)^{-1} - \left(\frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top\right)^{-1}\\ &= \left(\frac{1}{n}\mathbf{Z}_{2i}\mathbf{Z}_{2i}^\top - \frac{1}{\theta_i T}\mathbf{X}_2\mathbf{X}_2^\top\right)^{-1} \left\{\frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top - \left(\frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top - \frac{1}{\theta_i T}\mathbf{X}_2\mathbf{X}_2^\top\right)\right\} \left(\frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top\right)^{-1}\\ &= \theta_i^{-1} \left(\frac{1}{n}\mathbf{Z}_{2i}\mathbf{Z}_{2i}^\top - \frac{1}{\theta_i T}\mathbf{X}_2\mathbf{X}_2^\top\right)^{-1} \left(\frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top\right) \left(\frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top\right)^{-1}\end{aligned} \tag{62}$$

and

$$\frac{1}{p-q-1} \mathrm{tr}\left(\frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top\right)^{-1} = \mathcal{S}_{\mathrm{MP}}(0) + \mathrm{O_p}\left(p^{-1}\right) = \frac{1}{1-y} + \mathrm{O_p}\left(p^{-1}\right), \tag{63}$$

where $\mathcal{S}_{\mathrm{MP}}$ denotes the Stieltjes transform of the Marcenko–Pastur law. Then we have that

$$\begin{aligned}\frac{1}{p-q}\theta_i \mathrm{E}\left\{\mathrm{tr}\left(\theta_i \frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top - \frac{1}{T}\mathbf{X}_2\mathbf{X}_2^\top\right)^{-1}\right\} &= \mathrm{E}\left\{\frac{1}{p-q}\mathrm{tr}\left(\frac{1}{n}\mathbf{Z}_{2k}\mathbf{Z}_{2k}^\top\right)^{-1} + \mathrm{O_{a.s.}}\left(\theta_i^{-1}\right)\right\}\\ &= \mathrm{E}\left\{\frac{1}{1-y} + \mathrm{O_{a.s.}}\left(\theta_i^{-1}\right) + \mathrm{O_p}\left(p^{-1}\right)\right\} \to \frac{1}{1-y}.\end{aligned}$$

By Lemma A.2. in [17], it holds that

$$\{\mathbf{I}_n - \mathbf{A}(\theta_i)\}(k,k) \to \frac{1}{1+y(1-y)^{-1}} = 1-y,$$

which implies

$$\omega_{\mathbf{I}_n-\mathbf{A}(\theta_i)} = \lim_{n\to\infty}\frac{1}{n}\sum_{1\le k\le n}\left[\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}(k,k)\right]^2 = (1-y)^2.$$

By the similar argument, we can obtain that

$$\omega_{\mathbf{I}_T+\mathbf{B}(\theta_i)} = \lim_{T\to\infty}\frac{1}{T}\sum_{1\le k\le T}\left[\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}(k,k)\right]^2 = 1.$$

Now we come to the calculation of $\beta_{\mathbf{I}_n-\mathbf{A}(\theta_i)}$ and $\beta_{\mathbf{I}_T+\mathbf{B}(\theta_i)}$. Since $\theta_i \to +\infty$ as n goes to infinity, we have

$$\begin{aligned}
&\lim_{n\to\infty}\int_{-\infty}^{\infty}\frac{\theta_i}{\theta_i-x}dF_n(x)=1,\ \lim_{n\to\infty}\int_{-\infty}^{\infty}\frac{\theta_i^2}{(\theta_i-x)^2}dF_n(x)=1,\\
&\lim_{T\to\infty}\int_{-\infty}^{\infty}\frac{x}{\theta_i-x}dF_n(x)=0,\ \lim_{T\to\infty}\int_{-\infty}^{\infty}\frac{x^2}{(\theta_i-x)^2}dF_n(x)=0.
\end{aligned}$$

Then these calculations lead to

$$\begin{aligned}
\beta_{\mathbf{I}_n-\mathbf{A}(\theta_i)} &= \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}\,\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}^2 = \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}\left\{\mathbf{I}_n-2\mathbf{A}(\theta_i)+\mathbf{A}^2(\theta_i)\right\}\\
&= 1-2\lim_{n\to\infty}\left(\frac{p-q}{n}\int_{-\infty}^{\infty}\frac{\theta_i}{\theta_i-x}dF_n(x)\right)+\lim_{n\to\infty}\left\{\frac{p-q}{n}\int_{-\infty}^{\infty}\frac{\theta_i^2}{(\theta_i-x)^2}dF_n(x)\right\} = 1-2y+y=1-y,\\
\beta_{\mathbf{I}_T+\mathbf{B}(\theta_i)} &= \lim_{T\to\infty}\frac{1}{T}\mathrm{tr}\,\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}^2 = \lim_{T\to\infty}\frac{1}{T}\mathrm{tr}\left\{\mathbf{I}_T+2\mathbf{B}(\theta_i)+\mathbf{B}^2(\theta_i)\right\}\\
&= 1+2\lim_{T\to\infty}\left\{\frac{p-q}{T}\int_{-\infty}^{\infty}\frac{x}{\theta_i-x}dF_n(x)\right\}+\lim_{T\to\infty}\left\{\frac{p-q}{T}\int_{-\infty}^{\infty}\frac{x^2}{(\theta_i-x)^2}dF_n(x)\right\} = 1+0+0=1.
\end{aligned}$$

Thus, we can write

$$\begin{aligned}
\tilde{\sigma}_i^2 &= y\tilde{\sigma}_{i\mathbf{A}}^2 + c(1-y)^2\tilde{\sigma}_{i\mathbf{B}}^2 = y\{\omega_{\mathbf{I}_n-\mathbf{A}(\theta_i)}(\nu_i-3)+2\beta_{\mathbf{I}_n-\mathbf{A}(\theta_i)}\} + c(1-y)^2\{\omega_{\mathbf{I}_T+\mathbf{B}(\theta_i)}(\nu_i-3)+2\beta_{\mathbf{I}_T+\mathbf{B}(\theta_i)}\}\\
&= (y+c)(1-y)^2\nu_i - y(1-y)(1-3y) - c(1-y)^2.
\end{aligned}$$

Thus the proof is completed. □

Proof of Lemma 3.3. By the definition of $\mathbf{\Theta}_{1n}$ in (43) again, we know

$$\mathbf{\Theta}_{1n} = (1+\delta_i)^2\frac{1}{n}\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top - \theta_i^{-1}\frac{1}{T}\mathbf{X}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{X}_1^\top + \frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top + \frac{1}{T}\mathbf{X}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top - \mathrm{E}(\cdot), \tag{64}$$

where $\mathrm{E}(\cdot)$ is the expectation of all the preceding terms.

Denote

$$\begin{aligned}
&\boldsymbol{\eta}_{n1} = \frac{1}{n}\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top - \mathrm{E}\left[\frac{1}{n}\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right],\ \boldsymbol{\eta}_{n2} = \frac{1}{T}\mathbf{Y}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{Y}_1^\top - \mathrm{E}\left[\frac{1}{T}\mathbf{Y}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{Y}_1^\top\right],\\
&\boldsymbol{\eta}_{n3} = \sqrt{\theta_i}\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{Y}_1^\top - \mathrm{E}\left\{\sqrt{\theta_i}\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{Y}_1^\top\right\},\ \boldsymbol{\eta}_{n4} = \sqrt{\theta_i}\frac{1}{T}\mathbf{Y}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top - \mathrm{E}\left\{\sqrt{\theta_i}\frac{1}{T}\mathbf{Y}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top\right\}.
\end{aligned}$$

By the fact that $\mathbf{X}_1 = \mathbf{U}^\top\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{U}\mathbf{Y}_1$, we can write

$$\mathbf{U}\mathbf{\Theta}_{1n}\mathbf{U}^\top := \sum_{i=1}^{4}\mathbf{V}_{ni}, \tag{65}$$

where

$$\mathbf{V}_{n1} = (1+\delta_i)^2\mathbf{U}\left(\frac{1}{n}\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top - \mathrm{E}\left[\frac{1}{n}\mathbf{Z}_1\{\mathbf{I}_n-\mathbf{A}(\theta_i)\}\mathbf{Z}_1^\top\right]\right)\mathbf{U}^\top = (1+\delta_i)^2\mathbf{U}\boldsymbol{\eta}_{n1}\mathbf{U}^\top, \tag{66}$$

$$\begin{aligned}
\mathbf{V}_{n2} &= -\theta_i^{-1}\mathbf{U}\left(\frac{1}{T}\mathbf{X}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{X}_1^\top - \mathrm{E}\left[\frac{1}{T}\mathbf{X}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{X}_1^\top\right]\right)\mathbf{U}^\top\\
&= -\theta_i^{-1}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}\left(\frac{1}{T}\mathbf{Y}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{Y}_1^\top - \mathrm{E}\left[\frac{1}{T}\mathbf{Y}_1\{\mathbf{I}_T+\mathbf{B}(\theta_i)\}\mathbf{Y}_1^\top\right]\right)\mathbf{U}^\top\mathbf{\Lambda}^{\frac{1}{2}} = -\theta_i^{-1}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}\boldsymbol{\eta}_{n2}\mathbf{U}^\top\mathbf{\Lambda}^{\frac{1}{2}},
\end{aligned} \tag{67}$$

$$\mathbf{V}_{n3}=\mathbf{U}\left[\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top-\mathrm{E}\left\{\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top\right\}\right]\mathbf{U}^\top=\mathbf{U}\left[\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{Y}_1^\top-\mathrm{E}\left\{\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{Y}_1^\top\right\}\right]\mathbf{U}^\top\mathbf{\Lambda}^{\frac{1}{2}}$$
$$=\mathbf{U}\left[\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{Y}_1^\top-\mathrm{E}\left\{\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{Y}_1^\top\right\}\right]\mathbf{U}^\top\mathbf{\Lambda}^{\frac{1}{2}}=\theta_i^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\eta}_{n3}\mathbf{U}^\top\mathbf{\Lambda}^{\frac{1}{2}} \tag{68}$$
$$\mathbf{V}_{n4}=\mathbf{U}\left[\frac{1}{T}\mathbf{X}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top-\mathrm{E}\left\{\frac{1}{T}\mathbf{X}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top\right\}\right]\mathbf{U}^\top=\theta_i^{-\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}\boldsymbol{\eta}_{n4}\mathbf{U}^\top. \tag{69}$$

Similarly as the arguments in the proof of Lemma 3.2, it holds that, for $1\le j_1,j_2\le q$,

$$\mathbf{e}_{j_1}^\top\boldsymbol{\eta}_{n1}\mathbf{e}_{j_2}=\mathrm{O_p}\left(\frac{1}{\sqrt{n}}\right),\quad \mathbf{e}_{j_1}^\top\boldsymbol{\eta}_{n2}\mathbf{e}_{j_2}=\mathrm{O_p}\left(\frac{1}{\sqrt{n}}\right),\quad \mathbf{e}_{j_1}^\top\boldsymbol{\eta}_{n3}\mathbf{e}_{j_2}=\mathrm{O_p}\left(\frac{1}{\sqrt{n\lambda_i}}\right),\quad \mathbf{e}_{j_1}^\top\boldsymbol{\eta}_{n4}\mathbf{e}_{j_2}=\mathrm{O_p}\left(\frac{1}{\sqrt{n\lambda_i}}\right).$$

Noting that $\mathbf{U}$ is an orthogonal matrix, we have that

$$\mathbf{e}_{j_1}^\top\mathbf{V}_{n1}\mathbf{e}_{j_2}=\mathbf{e}_{j_1}^\top(1+\delta_i)^2\mathbf{U}\boldsymbol{\eta}_{n1}\mathbf{U}^\top\mathbf{e}_{j_2}=\mathrm{O_p}\left(\frac{1}{\sqrt{n}}\right),\ \mathbf{e}_{j_1}^\top\mathbf{V}_{n2}\mathbf{e}_{j_2}=-\mathbf{e}_{j_1}^\top\theta_i^{-1}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}\boldsymbol{\eta}_{n2}\mathbf{U}^\top\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{e}_{j_2}=\lambda_i^{-1}\lambda_{j_1}^{\frac{1}{2}}\lambda_{j_2}^{\frac{1}{2}}\mathrm{O_p}\left(\frac{1}{\sqrt{n}}\right),$$
$$\mathbf{e}_{j_1}^\top\mathbf{V}_{n3}\mathbf{e}_{j_2}=\mathbf{e}_{j_1}^\top\theta_i^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\eta}_{n3}\mathbf{U}^\top\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{e}_{j_2}=\lambda_i^{-1}\lambda_{j_2}^{\frac{1}{2}}\mathrm{O_p}\left(\frac{1}{\sqrt{n}}\right),\ \mathbf{e}_{j_1}^\top\mathbf{V}_{n4}\mathbf{e}_{j_2}=\mathbf{e}_{j_1}^\top\theta_i^{-\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}\boldsymbol{\eta}_{n4}\mathbf{U}^\top\mathbf{e}_{j_2}=\lambda_i^{-1}\lambda_{j_1}^{\frac{1}{2}}\mathrm{O_p}\left(\frac{1}{\sqrt{n}}\right).$$

Then by Chebyshev's inequality, we can deduce that

$$\|\mathbf{V}_{n1}\|_\infty=\mathrm{O_p}\left(\frac{q}{\sqrt{n}}\right),\quad \|\mathbf{V}_{n2}\|_\infty=\mathrm{O_p}\left(\frac{\sum_j\lambda_j}{\sqrt{n}\lambda_i}\right),\quad \|\mathbf{V}_{n3}+\mathbf{V}_{n4}\|_\infty=\mathrm{O_p}\left(\frac{\sqrt{q\sum_j\lambda_j}}{\sqrt{n}\lambda_i}\right),$$

where $\sqrt{q\sum_j\lambda_j}=\mathrm{o}(\sum_j\lambda_j)$. Thus we complete the proof by (65). □

Proof of Lemma 3.4. Recall the definition of $\mathbf{\Theta}_{2n}$ in (44):

$$\begin{aligned}\mathbf{\Theta}_{2n}=&(1+\delta_i)\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top-(1+\delta_i)\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_1^\top\\&+\frac{1}{\hat\lambda_i\theta_i}\frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top\mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top-\frac{(1+\delta_i)}{\hat\lambda_i}\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top\\&+\frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top-\frac{(1+\delta_i)}{\hat\lambda_i}\frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top\mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top+\frac{1}{T}\mathbf{X}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top.\end{aligned}$$

Noting that

$$\mathbf{M}^{-1}(\hat\lambda_i)-\mathbf{M}^{-1}(\theta_i)=\mathbf{M}^{-1}(\hat\lambda_i)\left\{\mathbf{M}(\theta_i)-\mathbf{M}(\hat\lambda_i)\right\}\mathbf{M}^{-1}(\theta_i)=-\frac{\delta_i}{\hat\lambda_i}\mathbf{M}^{-1}(\hat\lambda_i)\mathbf{F}_0\mathbf{M}^{-1}(\theta_i),$$

we decompose the first term in $\mathbf{\Theta}_{2n}$ as

$$\begin{aligned}&\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top\\&=\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\left\{\mathbf{M}^{-1}(\hat\lambda_i)-\mathbf{M}^{-1}(\theta_i)\right\}\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top+\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-2}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top\\&=-\frac{\delta_i}{\hat\lambda_i}\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat\lambda_i)\mathbf{F}_0\mathbf{M}^{-2}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top+\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-2}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top.\end{aligned}$$

On one hand, similar to the arguments in the proof of Theorem 2.1, we can derive that

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top\frac{\delta_i}{\hat\lambda_i}\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat\lambda_i)\mathbf{F}_0\mathbf{M}^{-2}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top\mathbf{e}_j\right|=\mathrm{O_p}\left(\frac{\sqrt{q}\delta_i}{\lambda_i}\right),$$
$$\max_{1\le j_1\ne j_2\le q}\left|\mathbf{e}_{j_1}^\top\frac{\delta_i}{\hat\lambda_i}\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top\mathbf{M}^{-1}(\hat\lambda_i)\mathbf{F}_0\mathbf{M}^{-2}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top\mathbf{e}_{j_2}\right|=\mathrm{O_p}\left(\frac{q\delta_i}{\lambda_i}\right).$$

On the other hand, similar to the proof of Lemma 3.2, we can get that

$$\frac{1}{n}\left(\mathbf{e}_j^\top \mathbf{Z}_1\mathbf{Z}_2^\top \mathbf{M}^{-2}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top \mathbf{e}_j - \mathrm{E}\left[\mathrm{tr}\left\{\mathbf{M}^{-2}(\theta_i)\right\}\right]\right) = \mathrm{O_p}\left(\frac{1}{\sqrt{n}}\right),$$

$$\frac{1}{n}\mathbf{e}_{j_1}^\top \mathbf{Z}_1\mathbf{Z}_2^\top \mathbf{M}^{-2}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top \mathbf{e}_{j_2} = \mathrm{O_p}\left(\frac{1}{\sqrt{n}}\right),$$

where $\frac{1}{n}\mathrm{E}\left\{\mathrm{tr}\mathbf{M}^{-2}(\theta_i)\right\} \to y$. It follows that

$$\max_{1\le j\le q}\left|\frac{1}{n}\mathbf{e}_j^\top (1+\delta_i)\mathbf{Z}_1\mathbf{Z}_2^\top \mathbf{M}^{-2}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top \mathbf{e}_j - \frac{1}{n}\mathrm{E}\left[\mathrm{tr}\left\{\mathbf{M}^{-2}(\theta_i)\right\}\right]\right| = \mathrm{O_p}\left(\frac{\sqrt{q}}{\sqrt{n}}\right)$$

$$\max_{1\le j_1\ne j_2\le q}\left|\frac{1}{n}\mathbf{e}_{j_1}^\top (1+\delta_i)\mathbf{Z}_1\mathbf{Z}_2^\top \mathbf{M}^{-2}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top \mathbf{e}_{j_2}\right| = \mathrm{O_p}\left(\frac{q}{\sqrt{n}}\right).$$

Similarly, we can get the following for other terms:

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top (1+\delta_i)\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_1^\top \mathbf{e}_j - 1\right| = \mathrm{O_p}\left(\frac{\sqrt{q}}{\sqrt{n}}\right), \quad \max_{1\le j_1\ne j_2\le q}\left|\mathbf{e}_{j_1}^\top (1+\delta_i)\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_1^\top \mathbf{e}_{j_2}\right| = \mathrm{O_p}\left(\frac{q}{\sqrt{n}}\right),$$

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top \frac{1}{\hat\lambda_i\theta_i}\frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top \mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top \mathbf{e}_j\right| = \mathrm{O_p}\left(\frac{\sqrt{\sum_j\lambda_j^2}}{\lambda_i^2}\right),$$

$$\max_{1\le j_1\ne j_2\le q}\left|\mathbf{e}_{j_1}^\top \frac{1}{\hat\lambda_i\theta_i}\frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top \mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top \mathbf{e}_{j_2}\right| = \mathrm{O_p}\left(\frac{\sum_j\lambda_j}{\lambda_i^2}\right),$$

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top \frac{(1+\delta_i)}{\hat\lambda_i}\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top \mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top \mathbf{e}_j\right| = \mathrm{O_p}\left(\frac{\sqrt{\sum_j\lambda_j}}{\lambda_i}\right),$$

$$\max_{1\le j_1\ne j_2\le q}\left|\mathbf{e}_{j_1}^\top \frac{(1+\delta_i)}{\hat\lambda_i}\frac{1}{n}\mathbf{Z}_1\mathbf{Z}_2^\top \mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{T}\mathbf{X}_2\mathbf{X}_1^\top \mathbf{e}_{j_2}\right| = \mathrm{O_p}\left(\frac{\sqrt{q\sum_j\lambda_j}}{\lambda_i}\right),$$

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top \frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top \mathbf{e}_j\right| = \mathrm{O_p}\left(\frac{\sqrt{\sum_j\lambda_j}}{\sqrt{n}\lambda_i}\right), \quad \max_{1\le j_1\ne j_2\le q}\left|\mathbf{e}_{j_1}^\top \frac{1}{n}\mathbf{Z}_1\mathbf{C}(\theta_i)\mathbf{X}_1^\top \mathbf{e}_{j_2}\right| = \mathrm{O_p}\left(\frac{\sqrt{q\sum_j\lambda_j}}{\sqrt{n}\lambda_i}\right),$$

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top \frac{(1+\delta_i)}{\hat\lambda_i}\frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top \mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top \mathbf{e}_j\right| = \mathrm{O_p}\left(\frac{\sqrt{\sum_j\lambda_j}}{\lambda_i}\right),$$

$$\max_{1\le j_1\ne j_2\le q}\left|\mathbf{e}_{j_1}^\top \frac{(1+\delta_i)}{\hat\lambda_i}\frac{1}{T}\mathbf{X}_1\mathbf{X}_2^\top \mathbf{M}^{-1}(\hat\lambda_i)\mathbf{M}^{-1}(\theta_i)\left(\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_2^\top\right)^{-1}\frac{1}{n}\mathbf{Z}_2\mathbf{Z}_1^\top \mathbf{e}_{j_2}\right| = \mathrm{O_p}\left(\frac{\sqrt{q\sum_j\lambda_j}}{\lambda_i}\right),$$

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top \frac{1}{T}\mathbf{X}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top \mathbf{e}_j\right| = \mathrm{O_p}\left(\frac{\sqrt{\sum_j\lambda_j}}{\sqrt{n}\lambda_i}\right), \quad \max_{1\le j_1\ne j_2\le q}\left|\mathbf{e}_{j_1}^\top \frac{1}{T}\mathbf{X}_1\mathbf{D}(\theta_i)\mathbf{Z}_1^\top \mathbf{e}_{j_2}\right| = \mathrm{O_p}\left(\frac{\sqrt{q\sum_j\lambda_j}}{\sqrt{n}\lambda_i}\right).$$

Thus, all these inequalities lead to

$$\max_{1\le j\le q}\left|\mathbf{e}_j^\top \mathbf{\Theta}_{2n}\mathbf{e}_j - (y-1)\right| = \mathrm{O_p}\left(\frac{\sqrt{q}\delta_i}{\lambda_i} + \frac{\sqrt{q}}{\sqrt{n}} + \frac{\sqrt{\sum_j\lambda_j^2}}{\lambda_i^2} + \frac{\sqrt{\sum_j\lambda_j}}{\lambda_i}\right),$$

$$\max_{1\le j_1\ne j_2\le q}\left|\mathbf{e}_{j_1}^\top \mathbf{\Theta}_{2n}\mathbf{e}_{j_2}\right| = \mathrm{O_p}\left(\frac{q\delta_i}{\lambda_i} + \frac{q}{\sqrt{n}} + \frac{\sum_j\lambda_j}{\lambda_i^2} + \frac{\sqrt{q\sum_j\lambda_j}}{\lambda_i}\right).$$

The proof is completed. □

CRediT authorship contribution statement

Junshan Xie: Methodology, Formal analysis, Writing - original draft . **Yicheng Zeng:** Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Lixing Zhu:** Conceptualization, Supervision, Writing - review & editing, Project administration, Funding acquisition.

Acknowledgments

The authors gratefully acknowledge a grant from the University Grants Council of Hong Kong and a NSFC, China grant (NSFC11671042). Drs Xie and Zeng are co-first authors. Zeng and Zhu are in charge of all revisions. The authors thank Editor, Associate editor and two referees for their constructive comments and suggestions that led to an improvement of the early manuscript.

References

[1] Z. Bai, J.W. Silverstein, No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices, Ann. Probab. 26 (1998) 316–345.
[2] Z. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, second ed., in: Springer Series in Statistics, Springer, New York, 2010.
[3] Z. Bai, J. Yao, Central limit theorems for eigenvalues in a spiked population model, Ann. Inst. Henri Poincaré Probab. Stat. 44 (2008) 447–474.
[4] A. Bloemendal, A. Knowles, H.-T. Yau, J. Yin, On the principal components of sample covariance matrices, Probab. Theory Relat. Field 164 (2016) 459–552.
[5] T.T. Cai, X. Han, G. Pan, Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices, Ann. Statist. 48 (2020) 1255–1280.
[6] P. Dharmawansa, I.M. Johnstone, A. Onatski, Local asymptotic normality of the spectrum of high-dimensional spiked F-ratios, 2014, arXiv preprint arXiv:1411.3875.
[7] X. Han, G. Pan, Q. Yang, A unified matrix model including both CCA and f matrices in multivariate analysis: the largest eigenvalue and its applications, Bernoulli 24 (2018) 3447–3468.
[8] X. Han, G. Pan, B. Zhang, The Tracy–Widom law for the largest eigenvalue of F type matrices, Ann. Statist. 44 (2016) 1564–1592.
[9] R.A. Horn, C.R. Johnson, Matrix Analysis, second ed., Cambridge University Press, Cambridge, 2013.
[10] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2001) 295–327.
[11] N.E. Karoui, E. Purdom, The non-parametric bootstrap and spectral analysis in moderate and high-dimension, in: Proceedings of Machine Learning Research, Vol. 89, PMLR, 2019, pp. 2115–2124.
[12] D. Paul, A. Alexander, Random matrix theory in statistics: A review, J. Statist. Plann. Inference 150 (2014) 1–29.
[13] J.W. Silverstein, The limiting eigenvalue distribution of a multivariate F matrix, SIAM J. Math. Anal. 16 (1985) 641–646.
[14] J.W. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices, J. Multivariate Anal. 55 (1995) 331–339.
[15] J.W. Silverstein, Z. Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, J. Multivariate Anal. 54 (1995) 175–192.
[16] K.W. Wachter, The limiting empirical measure of multiple discriminant ratios, Ann. Statist. 8 (1980) 937–957.
[17] Q. Wang, J. Yao, Extreme eigenvalues of large-dimensional spiked Fisher matrices with application, Ann. Statist. 45 (2017) 415–460.
[18] J. Wishart, The generalised product moment distribution in samples from a normal multivariate population, Biometrika 20 (1928) 32–52.
[19] J. Yao, S. Zheng, Z. Bai, Large Sample Covariance Matrices and High-Dimensional Data Analysis, Cambridge University Press, New York, 2015.