J. Appl. Probab. 1–18 (2024)
*doi:*10.1017/jpr.2024.8

THE LIMITING SPECTRAL DISTRIBUTION OF LARGE RANDOM PERMUTATION MATRICES

JIANGHAO LI,* **
HUANCHAO ZHOU,* ***
ZHIDONG BAI,* **** AND
JIANG HU,* ***** *KLASMOE and Northeast Normal University*

Abstract

We explore the limiting spectral distribution of large-dimensional random permutation matrices, assuming the underlying population distribution possesses a general dependence structure. Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{C}^{m\times n}$ be an $m \times n$ data matrix after self-normalization (n samples and m features), where $\mathbf{x}_j = (x^*_{1j}, \ldots, x^*_{mj})^*$. Specifically, we generate a permutation matrix $\mathbf{X}_\pi$ by permuting the entries of $\mathbf{x}_j$ ($j = 1, \ldots, n$) and demonstrate that the empirical spectral distribution of $\mathbf{B}_n = (m/n)\mathbf{U}_n\mathbf{X}_\pi\mathbf{X}^*_\pi\mathbf{U}^*_n$ weakly converges to the generalized Marčenko–Pastur distribution with probability 1, where $\mathbf{U}_n$ is a sequence of $p \times m$ non-random complex matrices. The conditions we require are $p/n \to c > 0$ and $m/n \to \gamma > 0$.

Keywords: Random permutation matrices; Stieltjes transform; limiting spectral distribution

2020 Mathematics Subject Classification: Primary 60B20
Secondary 62H10; 60B12

1. Introduction

In multivariate statistical analysis, the sample covariance matrix is an extremely important statistic; see [1] for more details. Its spectral properties serve as a fundamental theoretical underpinning for *principal component analysis* [17]. In situations where the population size remains 'small' while the sample size becomes sufficiently large, the spectral properties are well understood. Classical probability outcomes reveal that the sample covariance matrix is a good approximation of the population one. However, when both sample and population sizes escalate indefinitely, this is not the case. Consequently, modern statistics urgently require a more robust theoretical framework, possibly with the constraint that their aspect ratio $c_n := p/n$ approaches a finite limit $c \in (0, \infty)$. In this context, the asymptotic spectral properties of the sample covariance matrix have garnered increasing attention in recent years, initiated

Received 24 November 2022; accepted 16 January 2024.
* Postal address: School of Mathematics and Statistics, Northeast Normal University, Changchun, China.
** Email address: lijh458@nenu.edu.cn
*** Email address: zhouhc782@nenu.edu.cn
**** Postal address: School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China. Email address: baizd@nenu.edu.cn
***** Email address: huj156@nenu.edu.cn

by [21] and further developed by [4, 18, 26, 29]. Extensive research has been carried out on models with a general population covariance; see, e.g., [22, 24, 25] for references.

To make the following discussions precise, we detail some notation.

Definition 1. Let $\mathbf{A}$ be a $p \times p$ Hermitian matrix, and denote its real eigenvalues by $\lambda_i(\mathbf{A})$, $i = 1, 2, \ldots, p$. Then, the empirical spectral distribution (ESD) of $\mathbf{A}$ is defined by $F^{\mathbf{A}}(x) = (1/p)\sum_{i=1}^{p} 1(\lambda_i(\mathbf{A}) \leq x)$, where $1(A)$ is the indicator function of an event A.

Definition 2. If $G(x)$ is a function of bounded variation on the real line, then its Stieltjes transform is defined by

$$s_G(z) = \int \frac{1}{x - z}\, \mathrm{d}G(x),$$

where $z = u + \mathrm{i}v \in \mathbb{C}^+$. Further, the Stieltjes transform of $F^{\mathbf{A}}(x)$, which is the ESD of a $p \times p$ Hermitian matrix $\mathbf{A}$, is given by

$$s_{F^{\mathbf{A}}}(z) = \int \frac{1}{x - z}\, \mathrm{d}F^{\mathbf{A}}(x) = \frac{1}{p}\operatorname{tr}(\mathbf{A} - z\mathbf{I}_p)^{-1},$$

where $\mathbf{I}_p$ is the identity matrix.

Remark 1. For more details on the Stieltjes transform, we refer the reader to [13, 23] and references therein.

We briefly review some relevant theoretical background here. Assume $\mathbf{w}_j = \boldsymbol{\Sigma}_n^{1/2}\mathbf{v}_j$, $1 \leq j \leq n$, where $\boldsymbol{\Sigma}_n^{1/2}$ is the square root of the non-negative definite Hermitian matrix $\boldsymbol{\Sigma}_n$, and $\mathbf{V}_n = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is a $p \times n$ matrix. Its elements are independent and identically distributed (i.i.d.) complex random variables with zero means and unit variances. When $p, n \to \infty$, $p/n \to c \in (0, \infty)$, the ESD $F^{\boldsymbol{\Sigma}_n}$ of $\boldsymbol{\Sigma}_n$ tends to a non-random probability distribution H, and the sequence $\boldsymbol{\Sigma}_n$ is bounded in spectral norm. Then, the limiting spectral distribution (LSD) of the sample covariance matrix has been well studied in [22, 24]. The theory states that with probability 1, the ESD $F^{\mathbf{S}_n}$ of the sample covariance matrix $\mathbf{S}_n = (1/n)\sum_j^n \mathbf{w}_j\mathbf{w}_j^* = (1/n)\boldsymbol{\Sigma}_n^{1/2}\mathbf{V}_n\mathbf{V}_n^*\boldsymbol{\Sigma}_n^{1/2}$ weakly converges to a non-random distribution F as $n \to \infty$. Here, $\mathbf{W}_n^*$ and $\mathbf{V}_n^*$ are the conjugate transpose of matrices $\mathbf{W}_n$ and $\mathbf{V}_n$, respectively. For each $z \in \mathbb{C}^+ = \{u + \mathrm{i}v \colon v > 0\}$, $s(z) = s_F(z)$ is the unique solution to the equation

$$s(z) = \int \frac{1}{t(1 - c - czs(z)) - z}\, \mathrm{d}H(t) \tag{1}$$

in the set $\{s(z) \in \mathbb{C} \colon -(1 - c)/z + cs(z) \in \mathbb{C}^+\}$, where $s_F(z)$ is defined as the Stieltjes transform of F. For more details, we refer the reader to [3, 10, 28] for references.

In the fields of biology and finance, high-dimensional data with very small sample sizes are common. The *permutation approach* is an elegant technique in non-parametric statistics dating back to *Fisher's permutation test* [12]. The basic consideration is that permuting the observations of a sample preserves the magnitude of the sample (i.e. keeps sample moments of all orders invariant) while weakening the dependence between observations. Therefore, by comparing the statistics based on the original sample with that based on the permuted sample, we can find whether the original sample contains the assumed information. The permutation approach has been widely used for recovering data permutations from noisy observations [16, 20], permutation parallel analysis for selecting the number of factors and principal components [7, 11], and so on. Given these considerations, the sample covariance matrix of permuted

https://doi.org/10.1017/jpr.2024.8 Published online by Cambridge University Press

samples may be of particular interest within the field of random matrix theory. In this paper, our exploration is centered around this compelling matter: the LSD of eigenvalues of random permutation matrices. An investigation into the limiting properties of large-dimensional random permutation matrices can instigate new challenges for classical statistical theory and propel the progression of large-dimensional random matrix theory

In what follows, we are ready to propose our model. We reprocess the raw data and divide the process into the following three steps. First, suppose that the raw data sample $\mathbf{z}_1, \ldots, \mathbf{z}_n \in \mathbb{C}^m$, where $\mathbf{z}_j = (z_{1j}^*, \ldots, z_{mj}^*)^*$, $j = 1, \ldots, n$. Then, we centralize the raw data sample. Let $\tilde{\mathbf{z}}_j = \mathbf{z}_j - \bar{z}_j \cdot \mathbf{1}$, where $\mathbf{1} = (1, \ldots, 1)^*$ denotes the m-dimensional vector of ones and $\bar{z}_j = (1/m) \sum_{i=1}^m z_{ij}$. Accordingly, $\tilde{z}_{ij} = z_{ij} - \bar{z}_j$. Second, we further standardize the sample by letting

$$x_{ij} = \frac{\tilde{z}_{ij}}{\sqrt{\sum_{i=1}^m |z_{ij} - \bar{z}_j|^2}} = \frac{z_{ij} - \bar{z}_j}{\sqrt{\sum_{i=1}^m |z_{ij} - \bar{z}_j|^2}}.$$

Thus, we could form the new $m \times n$ data matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{C}^{m\times n}$ (n samples and m features), where $\mathbf{x}_j = (x_{1j}^*, \ldots, x_{mj}^*)^*$. Finally, we focus on the self-normalized samples below. We shuffle each column of $\mathbf{X}$ independently by randomly permuting its entries. Each column has a random permutation independent of all the other columns, and the permutation $\boldsymbol{\pi}_j$ ($j = 1, \ldots, n$) permutes the entries of the jth column of $\mathbf{X}$, where $\pi_1, \ldots, \pi_n$ are n independent permutations of the set $\{1, \ldots, m\}$. Then, we obtain a permuted data matrix $\mathbf{X}_\pi$, so $\mathbf{X}_\pi$ has entries $(\mathbf{X}_\pi)_{ij} = x_{\pi_i^{(j)}, j}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. Let $\mathbf{x}_{\pi^{(1)}}, \mathbf{x}_{\pi^{(2)}}, \ldots, \mathbf{x}_{\pi^{(n)}}$ be the n columns of $\mathbf{X}_\pi$. We propose a random permutation matrix model

$$\mathbf{B}_n = \frac{m}{n} \mathbf{U}_n \mathbf{X}_\pi \mathbf{X}_\pi^* \mathbf{U}_n^*, \tag{2}$$

where $\mathbf{U}_n$ is a sequence of $p \times m$ non-random complex matrices.

In the literature, there are a few works on the eigenvalue distribution of random permutation matrices based on data permutation. The closest studies to this subject are the recent papers [14, 15], which prove that, with probability 1, the ESD of $(m/n)\mathbf{N}_\pi \mathbf{N}_\pi^\top$ for random permutation matrices $\mathbf{N}_\pi$ weakly converges to the generalized Marčenko–Pastur distribution. Here, $\mathbf{N} = \mathbf{UX}$, $\mathbf{U}$ is an $m \times m$ diagonal matrix, and $\mathbf{N}_\pi$ is the randomly row-permuted matrix of $\mathbf{N}$, i.e. for each $l \in \{1, \ldots, m\}$, the entries of the lth row of $\mathbf{N}$ are randomly permuted. Compared to the results in [14, 15], we examine the LSD of a covariance matrix comprised of column-permuted data with a non-permuted matrix $\mathbf{U}_n$. Moreover, we do not require $\mathbf{U}_n$ to be diagonal.

The approach used in this article is similar to that used by our third author in [5], i.e. the Stieltjes transform $s_{F^{\mathbf{B}_n}}(z)$ converges with probability 1 as $n \to \infty$ to a limit, which can be divided into the following three key steps. We first apply the martingale technique to obtain almost sure convergence of the random part. Secondly, we select a deterministic matrix $\mathbf{K}$ to ensure the convergence of the non-random part. Finally, the existence and uniqueness of the solution of the system of equations (1) is established.

Throughout the paper, let $|\mathbf{A}|$ be the matrix whose (i, j)th entry is the absolute value of the (i, j)th entry of $\mathbf{A}$. We denote the trace of a matrix $\mathbf{A}$ by $\operatorname{tr}(\mathbf{A})$. We denote the Hadamard product of two matrices $\mathbf{A}$ and $\mathbf{C}$ of the same size by $\mathbf{A} \circ \mathbf{C}$. For a $p \times n$ matrix $\mathbf{A}$, the notation $\|\mathbf{A}\|$ means the spectral norm of the matrix $\mathbf{A}$. The L_p norm of a vector $\mathbf{x} \in \mathbb{R}^m$ is defined as $\|\mathbf{x}\|_p = \left(\sum_{i=1}^m |x_i|^p\right)^{1/p}$, $p \geq 1$. The symbol $\xrightarrow{\text{a.s.}}$ means almost sure convergence, and '$*$' denotes the conventional conjugate transpose. We use big-O and little-o notation to express the asymptotic relationship between two quantities. And we also use L to denote various positive universal constants which may be different from one line to the next.

The remainder of the paper is structured as follows. In Section 2 we formulate the main result. In Section 3, application of this model is presented. Proofs of the main results are given in Section 4. Appendix A contains some auxiliary lemmas for convenience.

2. Main results

We begin to investigate the asymptotic properties of the ESD of random permutation matrices in high-dimensional frameworks, under the following assumptions.

Assumption 1. *$c_n = p/n \to c \in (0, \infty)$ and $\gamma_n = m/n \to \gamma \in (0, \infty)$ as $n \to \infty$.*

Assumption 2. *Let $\mathbf{U}_n$ be a sequence of $p \times m$ non-random complex matrices. The spectral norm of the matrix $\mathbf{T}_n = \mathbf{U}_n\mathbf{U}_n^*$ is uniformly bounded and its ESD converges weakly to a probability distribution H.*

Assumption 3. *Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{C}^{m\times n}$ be an $m \times n$ random matrix with, for all $j \in \{1, \ldots, n\}$, $\sum_{i=1}^m x_{ij} = 0$, $\sum_{i=1}^m |x_{ij}|^2 = 1$, and $\sum_{i=1}^m \mathbb{E}|x_{ij}|^4 \to 0$ uniformly in $j \le n$, as $m \to \infty$.*

The main result of this paper is as follows.

Theorem 1. *Suppose that Assumptions 1–3 hold. Then, almost surely, the ESD of $\mathbf{B}_n$ weakly converges to the generalized Marčenko–Pastur distribution, whose Stieltjes transform s(z) satisfies*

$$s(z) = \int \frac{1}{t(1 - c - czs(z)) - z} \, \mathrm{d}H(t), \qquad z \in \mathbb{C}^+. \tag{3}$$

Remark 2. Note that in this theorem we do not need the entries of $\mathbf{X}$ to be i.i.d. Moreover, the conditions $\sum_{i=1}^m x_{ij} = 0$ and $\sum_{i=1}^m |x_{ij}|^2 = 1$ hold after self-normalization and are necessary to make sure that the moment calculation is simpler under a very general dependence structure.

Remark 3. The condition $\sum_{i=1}^m \mathbb{E}|x_{ij}|^4 \to 0$ is straightforward when the fourth moment of the raw data is finite uniformly, by the relationship between uniformly integrable and moment.

Remark 4. When the entries x_{ij} of $\mathbf{X}$ are non-random, we apply the same shuffling process to each column of $\mathbf{X}$. Theorem 1 apparently still holds.

Specifically, considering a non-random normalized vector $\mathbf{x} = (x_1, x_2, \ldots, x_m)^\top$, we permute the entries of $\mathbf{x}$ randomly n times to obtain random permutation matrices $\mathbf{X}_\pi$. The permutation method is similar to sampling without replacement from finite populations. Thus, the following result is in line with Application 2.3 of the main theorem investigated by [5].

Corollary 1. *Let $\mathbf{X}_\pi = (\mathbf{X}_\pi)_{ij} \in \mathbb{R}^{m\times n}$ be a sequence of random permutation matrices with independent columns, where $\mathbf{X}_\pi$ is taken with respect to the non-random vector $\mathbf{x} = (x_1, x_2, \ldots, x_m)^\top$. Suppose that $n, m \to \infty$ such that $\gamma_n = m/n \to \gamma > 0$ and the vector $\mathbf{x}$ satisfies $\|\mathbf{x}\|_2^2 = 1$, $\|\mathbf{x}\|_4^4 \to 0$, and $\sum_{i=1}^m x_i = 0$. Then, with probability 1, the empirical spectral distribution of $\mathbf{X}_\pi\mathbf{X}_\pi^\top$ converges weakly to the Marčenko–Pastur distribution, whose Stieltjes transform $s = s(z)$ satisfies*

$$s = \frac{(1 - \gamma - z\gamma) + \sqrt{(1 - \gamma - z\gamma)^2 - 4\gamma^2 z}}{2z\gamma}, \qquad z \in \mathbb{C}^+.$$

https://doi.org/10.1017/jpr.2024.8 Published online by Cambridge University Press

3. Application of the model

In this section, we discuss a potential application of the introduced model. Assume that $W_1, \ldots, W_n$ are n observations from the AR(p) process defined as $W_t - \phi_1 W_{t-1} - \cdots - \phi_p W_{t-p} = Z_t$, where $\phi_1, \ldots, \phi_p$ are real constants, $(1 - \phi_1 z - \cdots - \phi_p z^p) \neq 0$ for $|z| \leqslant 1$, and $\{Z_t\} \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$, i.e. $\{Z_t\}$ are i.i.d. random variables with mean 0 and covariance $\sigma^2 > 0$. Recall that [8, 9] advocated a selection algorithm by employing a data-driven penalty to select the proper model among AR(p). The algorithm is defined as follows. First, choose a fixed integer K, which is believed to be greater than the true order p, and compute the Yule–Walker estimates $\hat{\phi}_{K1}, \ldots, \hat{\phi}_{KK}, \hat{\sigma}_K^2$ from the observed data, $\{W_t\}_{t=1}^n$. The residual sequence is given by $\hat{Z}_t = W_t - \hat{\phi}_{K1} W_{t-1} - \cdots - \hat{\phi}_{KK} W_{t-K}$ for $t = K+1, \ldots, n$. Centralize and standardize the residuals by subtracting the sample mean $\bar{Z}_t = 1/(n-K) \sum_{t=K+1}^m \widehat{Z}_t$ and standardized deviation $s_{nt} = \sqrt{1/(n-K-1) \sum_{t=K+1}^n (\widehat{Z}_t - \bar{Z}_t)^2}$. For simplicity, we use the same notation $\{\widehat{Z}_t\}_{K+1}^n$ for the normalized residuals. Then, choose a positive intege $K_1 \leq K$ for each $k = 0, \ldots, K_1$, select proper AR coefficients $\tilde{\phi}_{k1}, \ldots, \tilde{\phi}_{kk}$ and construct a pseudo-AR (k) series $Y_1^{(k)}, \ldots, Y_n^{(k)}$ from the model $Y_t^{(k)} - \tilde{\phi}_{k1} Y_{t-1}^{(k)} - \cdots - \tilde{\phi}_{kk} Y_{t-k}^{(k)} = \hat{Z}_t^*$, where $\{\hat{Z}_t^*\}$ is an i.i.d. sequence sampled from the normalized residuals, $\{\widehat{Z}_t\}_{K+1}^n$. The case $k = 0$ corresponds to $Y_t^{(0)} = \hat{Z}_t^*$. Note that $\tilde{\phi}_{k1}, \ldots, \tilde{\phi}_{kk}$ are user-selected coefficients in accordance with the real requirement of accuracy. Next, apply a larger set of pseudo-time series to approximate the range of correct penalty factors, and then ultimately make the estimated order of the model more accurate.

In fact, the data-driven selection criterion can easily extend to the causal ARMA (p, q) process satisfying the recursive equation $\Phi(B) W_t = \Theta(B) Z_t$, $\{Z_t\} \sim WN(0, \sigma^2)$, where the polynomial $\Phi(z) := 1 - \phi_1 z - \cdots - \phi_p z^p$ satisfies the causality condition $\Phi(z) \neq 0$ for $|z| \leqslant 1$, $\Theta(z) := 1 + \theta_1 z + \cdots + \theta_q z^q$, $WN(0, \sigma^2)$ stands for white noise with zero mean and variance σ^2, and B as usual denotes the backward shift operator. By [6, Theorem 3.1.1 (Causality Criterion), p. 85], the causal ARMA (p, q) process has the representation $W_t = \Phi^{-1}(B) \Theta(B) Z_t := \Psi(B) Z_t$. Thus, we have the desired representation $W_t = \sum_{i=0}^\infty \Psi_i Z_{t-i}$, where the sequence $\{\Psi_i\}$ is determined by $\Psi(z) = \sum_{i=0}^\infty \Psi_i z^i = \Theta(z)/\Phi(z)$, $|z| \leq 1$.

We deal with the causal ARMA (p, q) process similarly to the AR process. For the pair (K, Q), where it is believed that $K \geq p$ and $Q \geq q$, we could estimate the coefficients $\hat{\phi}_k$ and $\hat{\theta}_s$ from the observed data, $1 \leq k \leq K$, $1 \leq s \leq Q$. Then, centralize and normalize the residuals after computing them using the estimated coefficients. Note that each column is a time series, so the coefficients are estimated by column, and thus the residuals are also calculated by column. Next, choose positive integers $K_1 \leq K$, $Q_1 \leq Q$, for each pair (k, s) and construct the pseudo-ARMA (k, s) series $Y_{j,t}^{(k,s)}$ for $k = 1, \ldots, K_1$, $s = 1, \ldots, Q_1$ from the model

$$Y_{j,t}^{(k,s)} = \tilde{\Psi}(B) \hat{Z}_t^* = \tilde{\Psi}_{j,0} \hat{Z}_{j,t}^* + \tilde{\Psi}_{j,1} \hat{Z}_{j,t-1}^* + \tilde{\Psi}_{j,2} \hat{Z}_{j,t-2}^* + \cdots, \tag{4}$$

where the coefficients $\tilde{\Psi}(B)$ are determined by the preselected coefficients of $\Phi(B)$ and $\Theta(B)$ with the relation $\tilde{\Psi}(B) = \Phi^{-1}(B) \Theta(B)$, and set the residuals $\{\hat{Z}_t^*\}$ to be sampled from the normalized residuals. $\hat{Z}_{j,t}^*$ is assigned to be 0 if there is no such residual. Say, if the last coefficient is too small, the order of the model would be a lower one. Note that $\hat{\Psi}(B)$ is an infinite series and there are only finitely many residuals, so one has to truncate the backward operator $\hat{\Psi}(B)$ to a finite sum, say the first m terms. In that way, the jth reconstructed series

https://doi.org/10.1017/jpr.2024.8 Published online by Cambridge University Press

$\mathbf{Y}_j^{(k,s)} = (Y_{j,1}^{(k,s)}, \ldots, Y_{j,m}^{(k,s)})'$ in (4) can be written as

$$\mathbf{Y}_j^{(k,s)} = \tilde{\boldsymbol{\Psi}}\hat{\mathbf{Z}}_j^* \quad \text{for } k = 1, \ldots, K_1,\ s = 1, \ldots, Q_1, \tag{5}$$

where

$$\tilde{\boldsymbol{\Psi}} = \begin{pmatrix} \tilde{\Psi}_{j,0} & 0 & 0 & \cdots & 0 \\ \tilde{\Psi}_{j,1} & \tilde{\Psi}_{j,0} & 0 & \ldots & 0 \\ \tilde{\Psi}_{j,2} & \tilde{\Psi}_{j,1} & \tilde{\Psi}_{j,0} & \cdots & 0 \\ \vdots & \ldots & \ldots & \ddots & \vdots \\ \ldots & \ldots & \ldots & \ldots & \tilde{\Psi}_{j,0} \end{pmatrix}.$$

To avoid confusion, it is worth noting once more that the elements in the matrices $\tilde{\boldsymbol{\Psi}}$ are preselected here, similar to the way they were handled in [8, 9]. Inspired by (5), the sample from the normalized residuals can also be regarded as a procedure for permuting the normalized residual sequences. Thus, the construction of pseudo-time series may benefit from a permutation of the normalized residuals to reduce sampling variability. This approach introduces increased randomness and enables the construction of a larger set of pseudo-series. This, in turn, provides a sound rationale for our proposed model (2). That is, the columns of $\mathbf{X}_n$ are permuted (corresponding to the permuted residuals), while those of $\mathbf{U}_n$ (corresponding to the preselected coefficient matrices $\tilde{\boldsymbol{\Psi}}$) remain as is.

An attractive model order estimation technique was developed in [19, 27] based on the minimum eigenvalue of a covariance matrix derived from the observed data, which avoided the need for estimating the model parameters. In conjunction with the above discussion, we may make it feasible to estimate the true orders of the model by using the minimum eigenvalue of the covariance matrix of a larger set of pseudo-sequences. Consequently, the results established in this paper could offer useful insights for the theoretical analysis pertaining to data-driven model selection for order detection in time series. This prospect is indeed intriguing for future research. Nevertheless, the issue of model selection falls outside the scope of random matrix theory. For this reason, our discussion is limited to a straightforward description of the structure of the reconstructed models in this paper. The further investigation of the spectral properties of large-dimensional random permutation matrices in order determination for the causal ARMA (p, q) process will be left for future work.

4. Proof of the main result

Proof of Theorem 1. Throughout the proof, for any z we write $z = u + \mathrm{i}v$, where u, v are the real and imaginary parts of z. Since $z \in \mathbb{C}^+$, we always have $v > 0$. For convenience, in the following we write

$$\mathbf{B}_n := \frac{m}{n}\mathbf{U}_n\mathbf{X}_\pi\mathbf{X}_\pi^*\mathbf{U}_n^* = \frac{m}{n}\sum_{k=1}^{n}\mathbf{U}_n\mathbf{x}_{\pi(k)}\mathbf{x}_{\pi(k)}^*\mathbf{U}_n^*, \qquad \mathbf{B}_{k,n} := \left(\sum_{k=1}^{n}\mathbf{r}_k\mathbf{r}_k^*\right) - \mathbf{r}_k\mathbf{r}_k^*,$$

and $\mathbf{r}_k = \sqrt{m/n}\mathbf{U}_n\mathbf{x}_{\pi(k)}$, $k = 1, \ldots, n$. Note that the matrix $\mathbf{B}_{k,n}$ is obtained from $\mathbf{B}_n$ with $\mathbf{r}_k$ removed. Recall that $s_{F^{\mathbf{B}_n}}(z) = s_n(z) = p^{-1}\,\mathrm{tr}\,((m/n)\mathbf{U}_n\mathbf{X}_\pi\mathbf{X}_\pi^*\mathbf{U}_n^* - z\mathbf{I}_p)^{-1} = p^{-1}\,\mathrm{tr}\,\mathcal{B}_n^{-1}$, where

https://doi.org/10.1017/jpr.2024.8 Published online by Cambridge University Press

$\mathcal{B}_n := \mathbf{B}_n - z\mathbf{I}_p$. We also write $\mathcal{B}_{k,n} := \mathbf{B}_{k,n} - z\mathbf{I}_p$. With this observation, we can make the following moment calculations:

$$\mathbb{E}x_{\pi_i^{(1)},j} = \mathbb{E}\frac{\sum_{i=1}^m x_{ij}}{m} = 0,$$

$$\mathbb{E}|x_{\pi_i^{(1)},j}|^2 = \mathbb{E}\frac{\sum_{i=1}^m |x_{ij}|^2}{m} = \frac{1}{m},$$

$$\mathbb{E}x_{\pi_{i_1}^{(1)},j}\bar{x}_{\pi_{i_2}^{(1)},j} = \mathbb{E}\frac{\sum_{i_1\neq i_2}^m x_{i_1 j}\bar{x}_{i_2 j}}{m(m-1)} = \frac{-1}{m(m-1)}, \qquad i_1 \neq i_2.$$

We further write $\mathbf{\Sigma}_\pi := \mathbb{E}x_{\pi^{(j)}}x^*_{\pi^{(j)}}$ for $j = 1, \ldots, n$, which is in fact independent of j and will be denoted as $\mathbf{\Sigma}_\pi = (\sigma_{il})$ for later use. Write $\widetilde{\mathbf{\Sigma}}_\pi = \mathbf{I}_m - (1/m)\mathbf{1}\mathbf{1}^*$, which is in fact $(m-1)\mathbf{\Sigma}_\pi$, i.e. $\mathbf{\Sigma}_\pi = (1/(m-1))\widetilde{\mathbf{\Sigma}}_\pi$, and $\mathbf{1} = (1, \ldots, 1)^*$.

According to [3, Theorem B.9], the almost sure convergence of the Stieltjes transform $s_n(z)$ to $s(z)$ ensures that the ESD weakly converges to a probability distribution almost surely. Thus, we proceed with the proof by the following two steps:

(i) $s_n(z) - \mathbb{E}s_n(z) \xrightarrow{\text{a.s.}} 0$;

(ii) $\mathbb{E}s_n(z) \to s(z)$, which satisfies (3).

Step (i): $s_n(z) - \mathbb{E}s_n(z) \xrightarrow{\text{a.s.}} 0$. Let $\mathbb{E}_k(\,\cdot\,) = \mathbb{E}(\,\cdot \mid \mathbf{x}_{\pi^{(k+1)}}, \ldots, \mathbf{x}_{\pi^{(n)}})$ denote the conditional expectation with respect to the σ-field generated by $\mathbf{x}_{\pi^{(k+1)}}, \ldots, \mathbf{x}_{\pi^{(n)}}$. It follows that $\mathbb{E}s_n(z) = \mathbb{E}_n(s_n(z))$ and $s_n(z) = \mathbb{E}_0(s_n(z))$. Similar to [2], we can write $s_n(z) - \mathbb{E}s_n(z)$ as the sum of martingale differences, i.e.

$$\begin{aligned} s_n(z) - \mathbb{E}s_n(z) &= \mathbb{E}_0 s_n(z) - \mathbb{E}_n s_n(z) \\ &= \sum_{k=1}^n (\mathbb{E}_{k-1}s_n(z) - \mathbb{E}_k s_n(z)) \\ &= \frac{1}{p}\sum_{k=1}^n (\mathbb{E}_{k-1} - \mathbb{E}_k)\big(\operatorname{tr}\mathcal{B}_n^{-1} - \operatorname{tr}\mathcal{B}_{k,n}^{-1}\big) = \frac{1}{p}\sum_{k=1}^n (\mathbb{E}_{k-1} - \mathbb{E}_k)\varepsilon_k, \end{aligned}$$

where $\varepsilon_k := \operatorname{tr}\mathcal{B}_n^{-1} - \operatorname{tr}\mathcal{B}_{k,n}^{-1}$. Notice that

$$\mathcal{B}_n^{-1} = \mathcal{B}_{k,n}^{-1} - \frac{\mathcal{B}_{k,n}^{-1}(\mathbf{r}_k\mathbf{r}_k^*)\mathcal{B}_{k,n}^{-1}}{1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k}.$$

By Lemma 1, we have

$$|\varepsilon_k| \leq \frac{\|\mathcal{B}_{k,n}^{-1}\mathbf{r}_k\|^2}{|1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k|} \leq v^{-1}.$$

It follows that $(\mathbb{E}_{k-1} - \mathbb{E}_k)\varepsilon_k$ forms a bounded martingale difference sequence. By the Burkholder inequality (see Lemma 2), this yields

$$\mathbb{E}|s_n(z) - \mathbb{E}s_n(z)|^q \leq K_q p^{-q}\mathbb{E}\left(\sum_{k=1}^n |(\mathbb{E}_{k-1} - \mathbb{E}_k)\gamma_k|^2\right)^{q/2} \leq K_q\left(\frac{2}{v}\right)^q p^{-q/2}\left(\frac{p}{n}\right)^{-q/2},$$

https://doi.org/10.1017/jpr.2024.8 Published online by Cambridge University Press

which implies that it is summable for $q > 2$. By the Borel–Cantelli lemma, $s_n(z) - \mathbb{E}s_n(z) = p^{-1}\sum_{k=1}^{n}(\mathbb{E}_{k-1} - \mathbb{E}_k)\varepsilon_k \xrightarrow{\text{a.s.}} 0$ is obtained.

Step (ii): $\mathbb{E}s_n(z) \to s(z)$. The fundamental technique of the approach is to surmise the deterministic equivalent of $s_n(z)$ by writing it in the form $p^{-1}\operatorname{tr}(\mathbf{K} - z\mathbf{I}_p)^{-1}$ at first, where $\mathbf{K}$ is assumed to be deterministic. Then, it will be performed by taking the difference $p^{-1}\mathbb{E}\operatorname{tr}(\mathbf{B}_n - z\mathbf{I}_p)^{-1} - p^{-1}\mathbb{E}\operatorname{tr}(\mathbf{K} - z\mathbf{I}_p)^{-1}$ and, during the calculation, determining $\mathbf{K}$ such that the difference tends to zero. Define

$$\mathbf{K} = \mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi\mathbf{U}_n^*\left(1 + \frac{m}{n}\mathbb{E}\operatorname{tr}(\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi)\right)^{-1}.$$

Since

$$(\mathbf{B}_n - z\mathbf{I}_p) - (\mathbf{K} - z\mathbf{I}_p) = \left\{\sum_{i=1}^{n}\mathbf{r}_k\mathbf{r}_k^*\right\} - \mathbf{K},$$

by (20) of Lemma 3 and the resolvent identity $\mathbf{A}^{-1} - \mathbf{C}^{-1} = -\mathbf{A}^{-1}(\mathbf{A} - \mathbf{C})\mathbf{C}^{-1}$ for any $p \times p$ invertible matrices $\mathbf{A}$ and $\mathbf{C}$, we have

$$\begin{aligned}
&(\mathbf{K} - z\mathbf{I}_p)^{-1} - (\mathbf{B}_n - z\mathbf{I}_p)^{-1}\\
&\quad = (\mathbf{K} - z\mathbf{I}_p)^{-1}((\mathbf{B}_n - z\mathbf{I}_p) - (\mathbf{K} - z\mathbf{I}_p))(\mathbf{B}_n - z\mathbf{I}_p)^{-1}\\
&\quad = \left\{\sum_{k=1}^{n}(\mathbf{K} - z\mathbf{I}_p)^{-1}\mathbf{r}_k\mathbf{r}_k^*(\mathbf{B}_n - z\mathbf{I}_p)^{-1}\right\} - (\mathbf{K} - z\mathbf{I}_p)^{-1}\mathbf{K}(\mathbf{B}_n - z\mathbf{I}_p)^{-1}\\
&\quad = \left\{\sum_{k=1}^{n}\frac{(\mathbf{K} - z\mathbf{I}_p)^{-1}(\mathbf{r}_k\mathbf{r}_k^*)\mathcal{B}_{k,n}^{-1}}{1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k}\right\} - (\mathbf{K} - z\mathbf{I}_p)^{-1}\mathbf{K}(\mathbf{B}_n - z\mathbf{I}_p)^{-1}.
\end{aligned} \tag{6}$$

Let $\mathbf{T}_n^0 = \mathbf{I}_p$ and $\mathbf{T}_n^1 = \mathbf{T}_n$. Multiplying $(1/p)\mathbf{T}_n^l$ for $l = 0, 1$ on both sides of (6), and then taking the trace and expectation, we obtain that

$$\begin{aligned}
&\frac{1}{p}\mathbb{E}\operatorname{tr}\mathbf{T}_n^l(\mathbf{K} - z\mathbf{I}_p)^{-1} - \frac{1}{p}\mathbb{E}\operatorname{tr}\mathbf{T}_n^l(\mathbf{B}_n - z\mathbf{I}_p)^{-1}\\
&\quad = \frac{1}{p}\mathbb{E}\left\{\sum_{k=1}^{n}\frac{\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{T}_n^l(\mathbf{K} - z\mathbf{I}_p)^{-1}\mathbf{r}_k}{1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k}\right\} - \frac{1}{p}\mathbb{E}\operatorname{tr}\mathbf{T}_n^l(\mathbf{K} - z\mathbf{I}_p)^{-1}\mathbf{K}(\mathbf{B}_n - z\mathbf{I}_p)^{-1}.
\end{aligned} \tag{7}$$

Note that

$$\begin{aligned}
&\mathbb{E}\left|(1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k) - \left(1 + \frac{m}{n}\mathbb{E}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi\right)\right|^2\\
&\qquad \le \mathbb{E}\Bigg|\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k - \frac{m}{n}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_{k,n}^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi\\
&\qquad\quad + \frac{m}{n}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_{k,n}^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi - \frac{m}{n}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi\\
&\qquad\quad + \frac{m}{n}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi - \frac{m}{n}\mathbb{E}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi\Bigg|^2\\
&\qquad \le 3\gamma\Bigg(\mathbb{E}\left|\mathbf{x}_{\pi(k)}^*\mathbf{U}_n^*\mathcal{B}_{k,n}^{-1}\mathbf{U}_n\mathbf{x}_{\pi(k)} - \operatorname{tr}\mathbf{U}_n^*\mathcal{B}_{k,n}^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi\right|^2
\end{aligned} \tag{8}$$

$$+\mathbb{E}\left|\frac{1}{m-1}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_{k.n}^{-1}\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi-\frac{1}{m-1}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi\right|^2 \tag{9}$$

$$+\mathbb{E}\left|\frac{1}{m-1}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi-\frac{1}{m-1}\mathbb{E}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi\right|^2\Bigg), \tag{10}$$

where the last inequality follows from Jensen's inequality. For (8), writing $\mathbf{B}=\mathbf{U}_n^*\mathcal{B}_{k,n}^{-1}\mathbf{U}_n=(b_{il})$ and $\boldsymbol{\Sigma}_\pi=(\sigma_{il})$, we can obtain that

$$\mathbb{E}\left|\mathbf{x}_{\pi^{(k)}}^*\mathbf{U}_n^*\mathcal{B}_{k,n}^{-1}\mathbf{U}_n\mathbf{x}_{\pi^{(k)}}-\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_{k,n}^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi\right|^2=\mathbb{E}\left|\sum_{i=1}^m\sum_{l=1}^m b_{il}\big(x_{\pi_i^{(k)},k}x_{\pi_l^{(k)},k}-\sigma_{il}\big)\right|^2\to 0. \tag{11}$$

Since the calculation of (11) is tedious, we postpone it to the appendix for interested readers. For (9), it follows from Lemma 1 that, as $m\to\infty$,

$$\left|\frac{1}{m-1}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_{k.n}^{-1}\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi-\frac{1}{m-1}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi\right|^2\le\frac{\left\|\mathbf{U}_n^*\widetilde{\boldsymbol{\Sigma}}_\pi\mathbf{U}_n\right\|^2}{(m-1)^2v^2}\to 0,$$

which implies that (9) converges to zero. For (10), by the similar martingale decomposition in Step (i) and Lemma 2, we have, as $n\to\infty$,

$$\begin{aligned}&\mathbb{E}\left|\frac{1}{m-1}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi-\frac{1}{m-1}\mathbb{E}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi\right|^2\\&\le K_2(m-1)^{-2}\mathbb{E}\left(\sum_{k=1}^n\left|(\mathbb{E}_{k-1}-\mathbb{E}_k)(\operatorname{tr}\mathbf{U}_n^*\widetilde{\boldsymbol{\Sigma}}_\pi\mathbf{U}_n\mathcal{B}_n^{-1}-\operatorname{tr}\mathbf{U}_n^*\widetilde{\boldsymbol{\Sigma}}_\pi\mathbf{U}_n\mathcal{B}_{k,n}^{-1})\right|^2\right)\\&\le K_2\left(\frac{2\left\|\mathbf{U}_n^*\widetilde{\boldsymbol{\Sigma}}_\pi\mathbf{U}_n\right\|}{v}\right)^2(m-1)^{-1}\left(\frac{m-1}{n}\right)^{-1}\to 0.\end{aligned}$$

Therefore, by combining (8), (9), and (10), we obtain

$$\mathbb{E}\left|\big(1+\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k\big)-\left(1+\frac{m}{n}\mathbb{E}\operatorname{tr}\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi\right)\right|^2=o(1). \tag{12}$$

Moreover, note that

$$\left|\frac{1}{1+\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k}\right|\le\frac{|z|}{v}.$$

https://doi.org/10.1017/jpr.2024.8 Published online by Cambridge University Press

It then follows from (7) and (12) that

$$\begin{aligned}
&\frac{1}{p}\mathbb{E}\left\{\sum_{k=1}^{n}\frac{\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{r}_k}{1+\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k}\right\}-\frac{1}{p}\mathbb{E}\,\mathrm{tr}\,\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{K}(\mathbf{B}_n-z\mathbf{I}_p)^{-1}\\
&\qquad=\frac{1}{p}\sum_{k=1}^{n}\mathbb{E}\left\{\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{r}_k\left(1+\frac{m}{n}\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi\right)^{-1}\right\}\\
&\qquad\quad-\frac{1}{p}\mathbb{E}\,\mathrm{tr}\,\mathcal{B}_n^{-1}\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{K}+o(1)\\
&\qquad=\frac{1}{np}\sum_{k=1}^{n}\mathbb{E}\left\{\mathrm{tr}\,\mathcal{B}_{k,n}^{-1}\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi\mathbf{U}_n^*\left(1+\frac{m}{n}\mathbb{E}\,\mathrm{tr}\,\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi\mathbf{U}_n^*\right)^{-1}\right\}\\
&\qquad\quad-\frac{1}{p}\mathbb{E}\,\mathrm{tr}\,\mathcal{B}_n^{-1}\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{K}+o(1)\\
&\qquad=\frac{1}{np}\sum_{k=1}^{n}\mathbb{E}\{\mathrm{tr}\,\mathcal{B}_{k,n}^{-1}\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{K}\}-\frac{1}{p}\mathbb{E}\,\mathrm{tr}\,\mathcal{B}_n^{-1}\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{K}+o(1)\\
&\qquad=\frac{1}{np}\sum_{k=1}^{n}\{\mathbb{E}\,\mathrm{tr}\,\mathcal{B}_{k,n}^{-1}\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{K}-\mathbb{E}\,\mathrm{tr}\,\mathcal{B}_n^{-1}\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}\mathbf{K}\}+o(1)=o(1),
\end{aligned}$$

where for the last equality we used Lemma 1 again. Finally, we deduce that, for $l=0$, 1,

$$\frac{1}{p}\mathbb{E}\,\mathrm{tr}\,\mathbf{T}_n^l(\mathbf{K}-z\mathbf{I}_p)^{-1}-\frac{1}{p}\mathbb{E}\,\mathrm{tr}\,\mathbf{T}_n^l(\mathbf{B}_n-z\mathbf{I}_p)^{-1}\to 0. \tag{13}$$

Write

$$\begin{aligned}
(\mathbf{K}-z\mathbf{I}_p)^{-1}&=\left(\frac{\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi\mathbf{U}_n^*}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}-z\mathbf{I}_p\right)^{-1}\\
&=\left(\frac{\mathbf{T}_n}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}-z\mathbf{I}_p-\frac{(1/m)\mathbf{U}_n\mathbf{1}\mathbf{1}^*\mathbf{U}_n^*}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}\right)^{-1}.
\end{aligned}$$

For any real $\lambda>0$, it is easy to show that

$$\mathrm{Im}\left(\lambda+z(1+\frac{m}{n}\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi)\right)>v\qquad(z\in\mathbb{C}^+).$$

Then, we can check that

$$\begin{aligned}
\left\|\left(\frac{\mathbf{T}_n}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}-z\mathbf{I}_p\right)^{-1}\right\|&\le\max_{t\ge 0}\left|\frac{t}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}-z\right|^{-1}\\
&\le\max_{t\ge 0}\left|\frac{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}{t-z(1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi)}\right|\le L/v,
\end{aligned}$$

where the constant L may change from one appearance to the next. Consequently, we could find that

$$\frac{1}{p}\mathbb{E}\,\mathrm{tr}\left[\frac{(1/m)\mathbf{1}^*\mathbf{U}_n^*((\mathbf{T}_n)/(1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi)-z\mathbf{I}_p)^{-2}\mathbf{U}_n\mathbf{1}}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}\right]\to 0. \tag{14}$$

By using (21) from Lemma 3 and (14), we can show that

$$\frac{1}{p}\mathbb{E}\left\{\mathrm{tr}\left(\frac{\mathbf{T}_n}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}-z\mathbf{I}_p\right)^{-1}-\mathrm{tr}\,\mathcal{B}_n^{-1}\right\}\to 0. \tag{15}$$

Similarly, by using (21) from Lemma 3 again and Lemma 4, we have $\|(\mathbf{K}-z\mathbf{I}_p)^{-1}\|\le L$. Thus, we obtain

$$\left|\frac{1}{p}\frac{\mathbb{E}\,\mathrm{tr}\,(1/m)\mathbf{U}_n\mathbf{1}\mathbf{1}^*\mathbf{U}_n^*(\mathbf{K}-z\mathbf{I}_p)^{-1}}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}\right|\to 0. \tag{16}$$

Notice that

$$\left|\frac{1}{p}\mathbb{E}\,\mathrm{tr}\,\frac{1}{m}\mathbf{U}_n\mathbf{1}\mathbf{1}^*\mathbf{U}_n^*\mathcal{B}_n^{-1}\right|\to 0. \tag{17}$$

By (16) we obtain

$$1+\frac{z}{p}\mathbb{E}\left\{\mathrm{tr}\left(\frac{\mathbf{U}_n\widetilde{\boldsymbol{\Sigma}}_\pi\mathbf{U}_n^*}{1+(p/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}-z\mathbf{I}_p\right)^{-1}\right\}-\frac{(1/p)\mathbb{E}\,\mathrm{tr}\,\mathbf{T}_n\mathcal{B}_n^{-1}}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}\to 0$$

as $n\to\infty$. Then, combining (13), (15), and (17), it follows that

$$1+z\mathbb{E}s_n(z)-\frac{(1/p)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\widetilde{\boldsymbol{\Sigma}}_\pi\mathbf{U}_n\mathcal{B}_n^{-1}}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}\to 0,$$

where we utilized the identity relation between $\boldsymbol{\Sigma}_\pi$ and $\widetilde{\boldsymbol{\Sigma}}_\pi$. Consequently, we deduce that

$$1-c_n(1+z\mathbb{E}s_n(z))=\frac{1}{1+(m/n)\mathbb{E}\,\mathrm{tr}\,\mathbf{U}_n^*\mathcal{B}_n^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_\pi}+o(1).$$

Substituting this into (15), we obtain

$$\frac{1}{p}\mathbb{E}\{\mathrm{tr}\,(\mathbf{T}_n(1-c_n(1+z\mathbb{E}s_n(z)))-z\mathbf{I}_p)^{-1}\}-\mathbb{E}s_n(z)\to 0. \tag{18}$$

For any fixed $z\in\mathbb{C}^+$, $\mathbb{E}s_n(z)=p^{-1}\mathbb{E}\,\mathrm{tr}\,\mathcal{B}_n^{-1}$ is a bounded sequence. Thus, for any subsequence $\{n'\}$, there is a subsubsequence $\{n''\}$ such that $\mathbb{E}s_{n''}(z)$ converges to a limit $s(z)$. Then, from (18), $s(z)$ satisfies

$$s(z)=\int\frac{1}{t(1-c-czs(z))-z}\,\mathrm{d}H(t),\qquad z\in\mathbb{C}^+. \tag{19}$$

In [22], it is proved that for any $z\in\mathbb{C}^+$ (19) has a unique solution in $\mathbb{C}^+$. Thus, we conclude that $\mathbb{E}s_n(z)$ tends to a unique $s(z)$. We have therefore finished the proof of Theorem 1. □

https://doi.org/10.1017/jpr.2024.8 Published online by Cambridge University Press

Appendix A. Supporting results

In this appendix we list some results that are needed in the proof.

Lemma 1. ([24, Lemma 2.6].) *Let* $\mathbf{A}$ *and* $\mathbf{C}$ *be* $p \times p$ *Hermitian matrices. For* $\tau \in \mathbb{R}$, $q \in \mathbb{C}^p$, *and* $z = u + \mathrm{i}v \in \mathbb{C}^+$,

$$|\operatorname{tr}((\mathbf{C} - z\mathbf{I}_n)^{-1} - (\mathbf{C} + \tau\mathbf{q}\mathbf{q}^* - z\mathbf{I}_n)^{-1})\mathbf{A}| \le \frac{\|\mathbf{A}\|}{v}.$$

Lemma 2. (Burkholder inequality in [3, Lemma 2.12].) *Let* $\{\mathbf{X}_k\}$ *be a complex martingale difference sequence with respect to the increasing* σ*-field* $\{F_k\}$. *Then, for* $q > 1$,

$$\mathbb{E}\left|\sum \mathbf{X}_k\right|^q \le K_q \mathbb{E}\left(\sum |\mathbf{X}_k|^2\right)^{q/2}.$$

Lemma 3. *For any* $z \in \mathbb{C}$ *with* $\mathrm{Im}(z) \neq 0$, *any Hermitian matrix* $\mathbf{C} \in \mathbb{C}^{p\times p}$, *and* $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{C}^p$ *where* $\mathbf{C} + \boldsymbol{\alpha}\boldsymbol{\alpha}^* - z\mathbf{I}_p$, $\mathbf{C} + \boldsymbol{\alpha}\boldsymbol{\beta}^* - z\mathbf{I}_p$, *and* $\mathbf{C} - z\mathbf{I}_p$ *are invertible, we have*

$$\boldsymbol{\alpha}^*(\mathbf{C} + \boldsymbol{\alpha}\boldsymbol{\alpha}^* - z\mathbf{I}_p)^{-1} = \frac{\boldsymbol{\alpha}^*(\mathbf{C} - z\mathbf{I}_p)^{-1}}{1 + \boldsymbol{\alpha}^*(\mathbf{C} - z\mathbf{I}_p)^{-1}\boldsymbol{\alpha}}, \tag{20}$$

$$(\mathbf{C} + \boldsymbol{\alpha}\boldsymbol{\beta}^* - z\mathbf{I}_p)^{-1} - (\mathbf{C} - z\mathbf{I}_p)^{-1} = -\frac{(\mathbf{C} - z\mathbf{I}_p)^{-1}\boldsymbol{\alpha}\boldsymbol{\beta}^*(\mathbf{C} - z\mathbf{I}_p)^{-1}}{1 + \boldsymbol{\beta}^*(\mathbf{C} - z\mathbf{I}_p)^{-1}\boldsymbol{\alpha}}. \tag{21}$$

The formula in (20) can be regarded as a special case of [24, (2.2)], and (21) is a direct result of the resolvent identity

$$(\mathbf{A} + \mathbf{p}\mathbf{q}^*)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{p}\mathbf{q}^*\mathbf{A}^{-1}}{1 + \mathbf{q}^*\mathbf{A}^{-1}\mathbf{p}},$$

where $\mathbf{A}$ is a $p \times p$ Hermitian matrix and $\mathbf{p}, \mathbf{q} \in \mathbb{C}^p$.

Lemma 4. (Weyl's inequality) *Let* $\mathbf{A}$ *and* $\mathbf{C}$ *be* $p \times p$ *Hermitian matrices. Let* $\lambda_1(\mathbf{C}) \ge \lambda_2(\mathbf{C}) \ge \cdots \ge \lambda_p(\mathbf{C})$ *denote the* p *eigenvalues of* $\mathbf{C}$, *and* $\lambda_i(\mathbf{A})$ *the* i*th eigenvalue of* $\mathbf{A}$. *Then* $\lambda_i(\mathbf{A}) + \lambda_p(\mathbf{C}) \le \lambda_i(\mathbf{A} + \mathbf{C}) \le \lambda_i(\mathbf{A}) + \lambda_1(\mathbf{C})$.

Proof of (11). Before proceeding, let us introduce some notation. Let $\mathbf{e}_k$ be the $m \times 1$ vector with the kth element being 1 and the others zero. In addition, we will use Assumption 3 repeatedly, often without mention. For $k = 1, \ldots, n$, let

$$\sigma_{il} = \begin{cases} \mathbb{E}x_{\pi_i^{(k)},k}\bar{x}_{\pi_l^{(k)},k} = \mathbb{E}\dfrac{\sum_{i\neq l}^m x_{ik}\bar{x}_{lk}}{m(m-1)} = \dfrac{-1}{m(m-1)}, & i \neq l; \\ \mathbb{E}x_{\pi_i^{(k)},k}\bar{x}_{\pi_i^{(k)},k} = \mathbb{E}\dfrac{\sum_{i=1}^m x_{ik}\bar{x}_{ik}}{m} = \dfrac{1}{m}, & i = l. \end{cases}$$

We start with the quantity

$$\mathbb{E}\left|\sum_{i=1}^m \sum_{l=1}^m b_{il}\big(\bar{x}_{\pi_i^{(k)},k} x_{\pi_l^{(k)},k} - \sigma_{il}\big)\right|^2 =$$

$$\sum_{i_1=1}^m \sum_{i_2=1}^m \sum_{l_1=1}^m \sum_{l_2=1}^m b_{i_1l_1}\bar{b}_{i_2l_2}\mathbb{E}\big(\bar{x}_{\pi_{i_1}^{(k)},k} x_{\pi_{l_1}^{(k)},k} - \sigma_{i_1l_1}\big)\big(x_{\pi_{i_2}^{(k)},k}\bar{x}_{\pi_{l_2}^{(k)},k} - \sigma_{i_2l_2}\big).$$

Note that the matrix $\mathbf{B} = \mathbf{U}_n^* \mathcal{B}_{k,n}^{-1} \mathbf{U}_n = (b_{il})$ is independent of $\mathbf{x}_{\pi^{(k)}}$, and its spectral norm is uniformly bounded. In this expression, the random variables with the same first index (i, l) are dependent. Thus there are now five cases, depending on how many distinct indices there are among them.

One unique index: $i_1 = i_2 = l_1 = l_2$. In this case, we can write the summation as

$$\mathrm{I} = \sum_{i_1=1}^{m} |b_{i_1 i_1}|^2 \cdot \mathbb{E}\big(\big|x_{\pi_{i_1}^{(k)},k}\big|^2 - \sigma_{i_1 i_1}\big)^2 = \sum_{i_1=1}^{m} |b_{i_1 i_1}|^2 \cdot \big(\mathbb{E}\big|x_{\pi_{i_1}^{(k)},k}\big|^4 - \sigma_{i_1 i_1}^2\big),$$

where

$$\mathbb{E}\big|x_{\pi_{i_1}^{(k)},k}\big|^4 - \sigma_{i_1 i_1}^2 = \frac{\mathbb{E}\sum_{i_1=1}^{m} |x_{i_1 k}|^4}{m} - \frac{1}{m^2} = -\frac{1}{m^2} + o(1/m).$$

Here we used the condition $\sum_{i=1}^{m} \mathbb{E}|x_{ik}|^4 \to 0$ $(k = 1, \ldots, n)$. Note that

$$\sum_{i_1}^{m} \big|b_{i_1 i_1}\big|^2 = \sum_{i_1=1}^{m} \big(\mathbf{e}_{i_1}^* \mathbf{B} \mathbf{e}_{i_1}\big) \left(\mathbf{e}_{i_1}^* \overline{\mathbf{B}} \mathbf{e}_{i_1}\right) \le m\, \|\mathbf{B}\| \cdot \|\overline{\mathbf{B}}\|.$$

Thus, we conclude that $\mathrm{I} = o(1)$.

Two distinct indices. $i_1 = i_2 = l_1 \neq l_2$ and $i_1 = i_2 = l_2 \neq l_1$ are symmetric cases; $i_1 = l_2 = l_1 \neq i_2$ and $l_1 = i_2 = l_2 \neq i_1$ are symmetric cases. Then we have that

$$\begin{aligned}\mathrm{II} = 2 &\sum_{i_1=i_2=l_2\neq l_1}^{m} b_{i_1 l_1} \overline{b}_{i_1 i_1} \cdot \mathbb{E}\big(\overline{x}_{\pi_{i_1}^{(k)},k} x_{\pi_{l_1}^{(k)},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi_{i_1}^{(k)},k} \overline{x}_{\pi_{i_1}^{(k)},k} - \sigma_{i_1 i_1}\big) \\ &+ 2 \sum_{l_1=i_2=l_2\neq i_1}^{m} b_{i_1 l_1} \overline{b}_{l_1 l_1} \cdot \mathbb{E}\big(\overline{x}_{\pi_{i_1}^{(k)},k} x_{\pi_{l_1}^{(k)},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi_{l_1}^{(k)},k} \overline{x}_{\pi_{l_1}^{(k)},k} - \sigma_{l_1 l_1}\big),\end{aligned}$$

where for $i_1 \neq l_1$ we have

$$\begin{aligned}\mathbb{E}\big(\overline{x}_{\pi_{i_1}^{(k)},k} x_{\pi_{l_1}^{(k)},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi_{l_1}^{(k)},k} \overline{x}_{\pi_{l_1}^{(k)},k} - \sigma_{l_1 l_1}\big) &= \mathbb{E}\big(\overline{x}_{\pi_{i_1}^{(k)},k} x_{\pi_{l_1}^{(k)},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi_{i_1}^{(k)},k} \overline{x}_{\pi_{i_1}^{(k)},k} - \sigma_{i_1 i_1}\big) \\ &= \mathbb{E} x_{\pi_{i_1}^{(k)},k} (\overline{x}_{\pi_{i_1}^{(k)},k})^2 x_{\pi_{l_1}^{(k)},k} - \frac{-1}{m^2(m-1)} \\ &= \frac{-\mathbb{E}\sum_{i_1=1}^{m} |x_{i_1 k}|^4}{m(m-1)} - \frac{-1}{m^2(m-1)} \\ &= \frac{1}{m^2(m-1)} + o(1/m^2).\end{aligned}$$

Note that

$$\left|\sum_{i_1\neq l_1} b_{i_1 l_1} \overline{b}_{l_1 l_1}\right| = \left|\sum_{i_1\neq l_1} b_{i_1 l_1} \overline{b}_{i_1 i_1}\right| = \left|\sum_{i_1=1}^{m} \overline{b}_{i_1 i_1} \cdot \mathbf{e}_{i_1}^* \mathbf{B}(\mathbf{1} - \mathbf{e}_{i_1})\right| \le m^2 \|\overline{\mathbf{B}}\| \cdot \|\mathbf{B}\|.$$

Thus, we obtain $\mathrm{II} = o(1)$.

Two distinct indices: $i_1=i_2\neq l_1=l_2$, $i_1=l_1\neq l_2=i_2$, and $i_1=l_2\neq i_2=l_1$. Here we could simplify the summation as $\mathrm{III}=\mathrm{III}_1+\mathrm{III}_2+\mathrm{III}_3$, where

$$\mathrm{III}_1=\sum_{i_1=i_2\neq l_1=l_2}^{m}|b_{i_1l_1}|^2\cdot\mathbb{E}\big(\bar{x}_{\pi_{i_1}^{(k)},k}x_{\pi_{l_1}^{(k)},k}-\sigma_{i_1l_1}\big)^2;$$

$$\mathrm{III}_2=\sum_{i_1=l_1\neq l_2=i_2}^{m}b_{i_1i_1}\bar{b}_{i_2i_2}\cdot\mathbb{E}\big(\bar{x}_{\pi_{i_1}^{(k)},k}x_{\pi_{i_1}^{(k)},k}-\sigma_{i_1i_1}\big)\big(\bar{x}_{\pi_{i_2}^{(k)},k}x_{\pi_{i_2}^{(k)},k}-\sigma_{i_2i_2});$$

$$\mathrm{III}_3=\sum_{i_1=l_2\neq i_2=l_1}^{m}b_{i_1i_2}\bar{b}_{i_2i_1}\cdot\mathbb{E}\big(\bar{x}_{\pi_{i_1}^{(k)},k}x_{\pi_{i_2}^{(k)},k}-\sigma_{i_1i_2}\big)\big(x_{\pi_{i_2}^{(k)},k}\bar{x}_{\pi_{i_1}^{(k)},k}-\sigma_{i_2i_1}\big).$$

When $i_1\neq i_2$, we can write

$$\begin{aligned}\mathbb{E}\big(x_{\pi_{i_1}^{(k)},k}\bar{x}_{\pi_{i_2}^{(k)},k}-\sigma_{i_1i_2}\big)^2&=\mathbb{E}x^2_{\pi_{i_1}^{(k)},k}\big(\bar{x}_{\pi_{i_2}^{(k)},k}\big)^2-\sigma^2_{i_1i_2}\\&=\frac{\mathbb{E}\sum_{i_1\neq i_2}^{p}x_{i_1k}^2(\bar{x}_{i_2k})^2}{m(m-1)}-\frac{1}{m^2(m-1)^2}=O(m^{-2});\end{aligned}$$

$$\begin{aligned}\mathbb{E}\big(x_{\pi_{i_1}^{(k)},k}\bar{x}_{\pi_{i_1}^{(k)},k}-\sigma_{i_1i_1}\big)\big(x_{\pi_{i_2}^{(k)},k}\bar{x}_{\pi_{i_2}^{(k)},k}-\sigma_{i_2i_2}\big)&=\mathbb{E}\big|x_{\pi_{i_1}^{(k)},k}\big|^2\big|x_{\pi_{i_2}^{(k)},k}\big|^2-\sigma_{i_1i_1}\sigma_{i_2i_2}\\&=\frac{\mathbb{E}\sum_{i_1\neq i_2}^{m}|x_{i_1k}|^2|x_{i_2k}|^2}{m(m-1)}-\frac{1}{m^2}\\&=\frac{\mathbb{E}\big(\sum_{i_1=1}^{m}|x_{i_1k}|^2\big)^2-\mathbb{E}\big(\sum_{i_1=1}^{m}|x_{i_1k}|^4\big)}{m(m-1)}-\frac{1}{m^2}\\&=o(1/m^2).\end{aligned}$$

Here, $\mathrm{III}_2=o(1)$ follows from the fact that $\mathbb{E}\sum_{i_1=1}^{m}|x_{i_1k}|^2=1$. As to the term $\mathbb{E}x^2_{\pi_{i_1}^{(k)},k}\big(\bar{x}_{\pi_{i_2}^{(k)},k}\big)^2$, we check that

$$\Big|\mathbb{E}\sum_{i_1,i_2}x_{i_1k}^2(\bar{x}_{i_2k})^2\Big|\leq\mathbb{E}\sum_{i_1,i_2}|x_{i_1k}^2||\bar{x}_{i_2k}|^2\leq 1. \tag{22}$$

Note that

$$\sum_{i_1\neq l_1}|b_{i_1l_1}|^2=\Big|\operatorname{tr}\mathbf{B}\mathbf{B}^*-\sum_{i_1=1}^{m}|b_{i_1i_1}|^2\Big|\leq Lm\|\overline{\mathbf{B}}\|\cdot\|\mathbf{B}\|,$$

$$\Big|\sum_{i_1\neq l_1}b_{i_1i_2}\bar{b}_{i_2i_1}\Big|=\Big|\operatorname{tr}\mathbf{B}\overline{\mathbf{B}}-\sum_{i=1}^{m}|b_{i_1i_1}|^2\Big|\leq Lm\|\overline{\mathbf{B}}\|\cdot\|\mathbf{B}\|.$$

Thus, as $m\to\infty$, we could obtain

$$\mathrm{III}_1+\mathrm{III}_3\leq\sum_{i_1\neq l_1}^{m}b_{i_1l_1}\bar{b}_{i_1l_1}\Big|\mathbb{E}x^2_{\pi_{i_1}^{(k)},k}\big(\bar{x}_{\pi_{l_1}^{(k)},k}\big)^2\Big|+\sum_{i_1\neq i_2}^{m}b_{i_1i_2}\bar{b}_{i_2i_1}\Big|\mathbb{E}x^2_{\pi_{i_1}^{(k)},k}\big(\bar{x}_{\pi_{i_2}^{(k)},k}\big)^2\Big|=o(1).$$

Eventually, we get $\mathrm{III}=o(1)$.

https://doi.org/10.1017/jpr.2024.8 Published online by Cambridge University Press

Three distinct indices. In this case, we divide it into six classes, writing $\mathcal{I}_1 = \{i_1 = i_2$, other indices different$\}$; $\mathcal{I}_2 = \{i_1 = l_1$, other indices different$\}$; $\mathcal{I}_3 = \{i_1 = l_2$, other indices different$\}$; $\mathcal{I}_4 = \{l_1 = i_2$, other indices different$\}$; $\mathcal{I}_5 = \{i_2 = l_2$, other indices different$\}$; and $\mathcal{I}_6 = \{l_1 = l_2$, other indices different$\}$. In what follows, we proceed to deal with the summation of three distinct indices, $\mathrm{IV} = \mathrm{IV}_1 + \mathrm{IV}_2 + \mathrm{IV}_3 + \mathrm{IV}_4 + \mathrm{IV}_5 + \mathrm{IV}_6$. Here,

$$\mathrm{IV}_1 = \sum_{l_1,i_1,i_2,l_2 \in \mathcal{I}_1}^{m} b_{i_1 l_1} \bar{b}_{i_1 l_2} \cdot \mathbb{E}\big(\bar{x}_{\pi^{(k)}_{i_1},k} x_{\pi^{(k)}_{l_1},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi^{(k)}_{i1},k} \bar{x}_{\pi^{(k)}_{l_2},k} - \sigma_{i_1 l_2}\big);$$

$$\mathrm{IV}_2 = \sum_{l_1,i_1,i_2,l_2 \in \mathcal{I}_2}^{m} b_{i_1 i_1} \bar{b}_{i_2 l_2} \cdot \mathbb{E}\big(\big|x_{\pi^{(k)}_{i_1},k}\big|^2 - \sigma_{i_1 i_1}\big)\big(x_{\pi^{(k)}_{i_2},k} \bar{x}_{\pi^{(k)}_{l_2},k} - \sigma_{i_2 l_2}\big);$$

$$\mathrm{IV}_3 = \sum_{l_1,i_1,i_2,l_2 \in \mathcal{I}_3}^{m} b_{i_1 l_1} \bar{b}_{i_2 i_1} \cdot \mathbb{E}\big(\bar{x}_{\pi^{(k)}_{i_1},k} x_{\pi^{(k)}_{l_1},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi^{(k)}_{i_2},k} \bar{x}_{\pi^{(k)}_{i_1},k} - \sigma_{i_2 i_1}\big);$$

$$\mathrm{IV}_4 = \sum_{l_1,i_1,i_2,l_2 \in \mathcal{I}_4}^{m} b_{i_1 l_1} \bar{b}_{l_1 l_2} \cdot \mathbb{E}\big(\bar{x}_{\pi^{(k)}_{i_1},k} x_{\pi^{(k)}_{l_1},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi^{(k)}_{l_1},k} \bar{x}_{\pi^{(k)}_{l_2},k} - \sigma_{l_1 l_2}\big);$$

$$\mathrm{IV}_5 = \sum_{l_1,i_1,i_2,l_2 \in \mathcal{I}_5}^{m} b_{i_1 l_1} \bar{b}_{i_2 i_2} \cdot \mathbb{E}\big(\bar{x}_{\pi^{(k)}_{i_1},k} x_{\pi^{(k)}_{l_1},k} - \sigma_{i_1 l_1}\big)\big(\big|x_{\pi^{(k)}_{i_2},k}\big|^2 - \sigma_{i_2 i_2}\big);$$

$$\mathrm{IV}_6 = \sum_{l_1,i_1,i_2,l_2 \in \mathcal{I}_6}^{m} b_{i_1 l_1} \bar{b}_{i_2 l_1} \cdot \mathbb{E}\big(\bar{x}_{\pi^{(k)}_{i_1},k} x_{\pi^{(k)}_{l_1},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi^{(k)}_{i_2},k} \bar{x}_{\pi^{(k)}_{l_1},k} - \sigma_{i_2 l_1}\big).$$

By carefully checking the expectations, we point out that the main components of the above expectations are as follows:

For all $l_1, i_1, i_2, l_2 \in \mathcal{I}_1$, we have

$$\begin{aligned}\mathbb{E}\big|x_{\pi^{(k)}_{i_1},k}\big|^2 x_{\pi^{(k)}_{l_1},k} \bar{x}_{\pi^{(k)}_{l_2},k} &= \frac{1}{m(m-1)(m-2)} \mathbb{E} \sum_{l_1,i_1,i_2,l_2 \in \mathcal{I}_1}^{m} |x_{i_1 k}|^2 x_{l_1 k} \bar{x}_{l_2 k} \\ &= -\frac{1}{m(m-1)(m-2)} \left[\mathbb{E} \sum_{i_1 \neq l_1}^{m} |x_{i_1 k}|^2 |x_{l_1 k}|^2 - \mathbb{E} \sum_{i_1}^{m} |x_{i_1 k}|^4\right] = O(m^{-3}),\end{aligned}$$

where we used the fact that $\mathbb{E}\sum_{i_1=1}^{m} |x_{i_1 k}|^2 = 1$.

For all $l_1, i_1, i_2, l_2 \in \mathcal{I}_3$, we similarly obtain

$$\begin{aligned}\mathbb{E}\big(\bar{x}_{\pi^{(k)}_{i_1},k}\big)^2 x_{\pi^{(k)}_{l_1},k} x_{\pi^{(k)}_{i_2},k} &= \frac{1}{m(m-1)(m-2)} \mathbb{E} \sum_{l_1,i_1,i_2,l_2 \in \mathcal{I}_3}^{m} (\bar{x}_{i_1 k})^2 x_{l_1 k} x_{i_2 k} \\ &= -\frac{1}{m(m-1)(m-2)} \left[\mathbb{E} \sum_{i_1 \neq l_1}^{m} (\bar{x}_{i_1 k})^2 x_{l_1 k}^2 - \mathbb{E} \sum_{i_1}^{m} |x_{i_1 k}|^4\right] = O(m^{-3}),\end{aligned}$$

where we used (22) again.

And then, for all $l_1, i_1, i_2, l_2 \in \mathcal{I}_4$, $\mathbb{E}\big(x_{\pi^{(k)}_{l_1},k}\big)^2 \bar{x}_{\pi^{(k)}_{l_2},k}\bar{x}_{\pi^{(k)}_{i_1},k} = O(m^{-3})$ is immediate. Note that $\mathrm{IV}_2 = \overline{\mathrm{IV}}_5 = o(1)$ since

$$\begin{aligned}
&\Bigg|\sum_{l_1,i_1,i_2,l_2\in\mathcal{I}_2}^{m} b_{i_1 i_1}\bar{b}_{i_2 l_2}\Bigg| \\
&\quad = \Bigg|\operatorname{tr}\mathbf{B}\cdot\mathbf{1}^*\overline{\mathbf{B}}\mathbf{1} - \sum_{i_1=l_1=i_2,l_2}^{m} b_{i_1 i_1}\bar{b}_{i_1 l_2} - \sum_{i_1=l_1=l_2,i_2}^{m} b_{i_1 i_1}\bar{b}_{i_2 i_1} + \sum_{i_1=l_1=i_2=l_2}^{m} b_{i_1 i_1}\bar{b}_{i_1 i_1}\Bigg| \\
&\quad = \Bigg|\operatorname{tr}\mathbf{B}\cdot\mathbf{1}^*\overline{\mathbf{B}}\mathbf{1} - \mathbf{1}^*(\mathbf{B}_{\mathrm{d}}\overline{\mathbf{B}})\mathbf{1} - \mathbf{1}^*(\mathbf{B}_{\mathrm{d}}\overline{\mathbf{B}})\mathbf{1} + \sum_{i=1}^{m}|b_{ii}|^2\Bigg| = O(m^2),
\end{aligned}$$

where $\mathbf{B}_{\mathrm{d}}$ is the diagonal matrix of the diagonal entries of $\mathbf{B}$. Then, we turn to analyze the remainder terms. It follows that

$$\begin{aligned}
&\Bigg|\sum_{l_1,i_1,i_2,l_2\in\mathcal{I}_1}^{m} b_{i_1 l_1}\bar{b}_{i_1 l_2}\Bigg| = \Bigg|\mathbf{1}^*\mathbf{B}\mathbf{B}^*\mathbf{1} - \mathbf{1}^*\mathbf{B}_{\mathrm{d}}\overline{\mathbf{B}}\mathbf{1} - \mathbf{1}\mathbf{B}\overline{\mathbf{B}}_{\mathrm{d}}\mathbf{1} + \sum_{i=1}^{m}|b_{ii}|^2\Bigg| = O(m^2), \\
&\Bigg|\sum_{l_1,i_1,i_2,l_2\in\mathcal{I}_3}^{m} b_{i_1 l_1}\bar{b}_{i_2 i_1}\Bigg| = \Bigg|\mathbf{1}^*\mathbf{B}\overline{\mathbf{B}}\mathbf{1} - \mathbf{1}^*\mathbf{B}\overline{\mathbf{B}}_{\mathrm{d}}\mathbf{1} - \mathbf{1}^*\mathbf{B}_{\mathrm{d}}\mathbf{B}^*\mathbf{1} + \sum_{i=1}^{m}|b_{ii}|^2\Bigg| = O(m^2).
\end{aligned}$$

Similarly, we can obtain $\big|\sum_{l_1,i_1,i_2,l_2\in\mathcal{I}_4}^{m} b_{i_1 l_1}\bar{b}_{i_2 i_1}\big| = O(m^2)$ and $\big|\sum_{l_1,i_1,i_2,l_2\in\mathcal{I}_6}^{m} b_{i_1 l_1}\bar{b}_{i_2 i_1}\big| = O(m^2)$. Hence, $\mathrm{IV}_1 + \mathrm{IV}_3 + \mathrm{IV}_4 + \mathrm{IV}_6 = O(m^{-1})$. Thus, we conclude that the overall terms $\mathrm{IV} = o(1)$.

Four distinct indices: $i_1 \neq i_2 \neq l_1 \neq l_2$. In this case, we can write the summation as

$$\mathrm{V} = \sum_{i_1\neq i_2\neq l_1\neq l_2}^{p} b_{i_1 l_1}\bar{b}_{i_2 l_2}\cdot\mathbb{E}\big(\bar{x}_{\pi^{(k)}_{i_1},k}x_{\pi^{(k)}_{l_1},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi^{(k)}_{i_2},k}\bar{x}_{\pi^{(k)}_{l_2},k} - \sigma_{i_2 l_2}\big).$$

For $i_1 \neq i_2 \neq l_1 \neq l_2$, we write

$$\begin{aligned}
\mathbb{E}\big(\bar{x}_{\pi^{(k)}_{i_1},k}x_{\pi^{(k)}_{l_1},k} - \sigma_{i_1 l_1}\big)\big(x_{\pi^{(k)}_{i_2},k}\bar{x}_{\pi^{(k)}_{l_2},k} - \sigma_{i_2 l_2}\big) &= \mathbb{E}\bar{x}_{\pi^{(k)}_{i1},k}x_{\pi^{(k)}_{i_2},k}x_{\pi^{(k)}_{l_1},k}\bar{x}_{\pi^{(k)}_{l_2},k} - \frac{1}{m^2(m-1)^2} \\
&= \frac{\mathbb{E}\sum_{i_1\neq i_2\neq l_1\neq l_2}^{m}\bar{x}_{i_1 k}x_{l_1 k}x_{l_2 k}\bar{x}_{i_2 k}}{m(m-1)(m-2)(m-3)} \\
&\quad - \frac{1}{m^2(m-1)^2} = O(m^{-4}).
\end{aligned}$$

https://doi.org/10.1017/jpr.2024.8 Published online by Cambridge University Press

Note that

$$\sum_{i_1\neq i_2\neq l_1\neq l_2} b_{i_1l_1}\bar{b}_{i_2l_2}$$

$$=\Bigg[\sum_{\{i_1,l_1,i_2,l_2\}}-\sum_{\{i_1=i_2\}}-\sum_{\{i_1=l_1\}}-\sum_{\{i_1=l_2\}}-\sum_{\{l_1=i_2\}}-\sum_{\{l_1=l_2\}}-\sum_{\{i_2=l_2\}}+\sum_{\{i_1=i_2=l_1\}}+\sum_{\{i_1=i_2=l_2\}}$$

$$+\sum_{\{i_2=l_1=l_2\}}+\sum_{\{i_1=l_1=l_2\}}+\sum_{\{i_1=i_2,l_1=l_2\}}+\sum_{\{i_1=l_1,i_2=l_2\}}+\sum_{\{i_1=l_2,i_2=l_1\}}-\sum_{\{i_1=i_2=l_1=l_2\}}\Bigg](b_{i_1l_1}\bar{b}_{i_2l_2})$$

$$=(\mathbf{1}^*\mathbf{B}\mathbf{1})(\mathbf{1}^*\overline{\mathbf{B}}\mathbf{1})-\operatorname{tr}\mathbf{B}\cdot\mathbf{1}^*\overline{\mathbf{B}}\mathbf{1}-\operatorname{tr}\overline{\mathbf{B}}\cdot\mathbf{1}^*\mathbf{B}\mathbf{1}-2\mathbf{1}^*\mathbf{B}\overline{\mathbf{B}}\mathbf{1}-2\mathbf{1}^*\mathbf{B}\mathbf{B}^*\mathbf{1}+$$

$$2\mathrm{Re}(\mathbf{1}^*\mathbf{B}_{\mathrm{d}}\mathbf{B}^*\mathbf{1})+2\mathrm{Re}(\mathbf{1}^*\mathbf{B}\overline{\mathbf{B}}_{\mathrm{d}}\mathbf{1})+|\operatorname{tr}\mathbf{B}|^2+\operatorname{tr}\mathbf{B}\mathbf{B}^*+\operatorname{tr}\mathbf{B}\overline{\mathbf{B}}-\sum_{i=1}^{m}|b_{ii}|^2$$

$$=O(m^2).$$

Here we used the results established in the previous calculation of the four cases. Now, it is easy to see that the summation term of the b_{il} part is at most of the order of m^2. However, the expectation terms are of the order of m^{-4}, since $\sum_{i=1}^m \mathbb{E}|x_{ik}|^4 \to 0$ $(k=1,\ldots,n)$. Hence, the term $\mathrm{V}=o(1)$.

Putting together the results from the five cases, we obtain the desired result:

$$\mathbb{E}\left|\mathbf{x}^*_{\pi^{(k)}}\mathbf{U}_n^*\mathcal{B}_{k,n}^{-1}\mathbf{U}_n\mathbf{x}_{\pi^{(k)}}-\operatorname{tr}\mathbf{U}_n^{1/2}\mathcal{B}_{k,n}^{-1}\mathbf{U}_n\boldsymbol{\Sigma}_{\boldsymbol{\pi}}\right|^2=o(1).$$

□

Acknowledgements

We thank the Editor, Associate Editor, and the two referees for their insightful comments. Jianghao Li wishes to thank Ms. Xinlei Li for her encouragement and discussion.

Funding information

Zhidong Bai was partially supported by NSFC Grants Nos. 12171198, 12271536 and Team Project of Jilin Provincial Department of Science and Technology No. 20210101147JC. Jiang Hu was partially supported by NSFC Grants Nos. 12292980, 12292982, 12171078, 12326606, National Key R&D Program of China No. 2020YFA0714102, and Fundamental Research Funds for the Central Universities No. 2412023YQ003.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edn. John Wiley, Hoboken, NJ.

[2] Bai, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9**, 611–677.

[3] BAI, Z. D. AND SILVERSTEIN, J. (2010). *Spectral Analysis of Large-Dimensional Random Matrices*, 2nd edn. Springer, New York.
[4] BAI, Z. D. AND YIN, Y. Q. (1988). Convergence to the semicircle law. *Ann. Prob.* **16**, 863–875.
[5] BAI, Z. D. AND ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18**, 425–442.
[6] BROCKWELL, P. AND DAVIS, R. (1991). *Time Series: Theory and Methods*, 2nd edn. Springer, New York.
[7] BUJA, A. AND EYUBOGLU, N. (1992). Remarks on parallel analysis. *Multivariate Behav. Res.* 27, 509–540.
[8] CHEN, C., DAVIS, R. AND BROCKWELL, P. (1996). Order determination for multivariate autoregressive processes using resampling methods. *J. Multivariate Anal.* **57**, 175–190.
[9] CHEN, C., DAVIS, R., BROCKWELL, P. AND BAI, Z. D. (1993). Order determination for autoregressive processes using resampling methods. *Statist. Sinica* **3**, 481–500.
[10] COUILLET, R. AND DEBBAH, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge University Press.
[11] DOBRIBAN, E. (2020). Permutation methods for factor analysis and PCA. *Ann. Statist.* **48**, 2824–2847.
[12] FISHER, R. (1935). *The Design of Experiments*. Hafner, New York.
[13] GERONIMO, J. AND HILL, T. (2003). Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform. *J. Approx. Theory* **121**, 54–60.
[14] HONG, D., SHENG, Y. AND DOBRIBAN, E. (2020). Selecting the number of components in pca via random signflips. Preprint, arXiv:2012.02985.
[15] HU, J., WANG, S., ZHANG, Y. AND ZHOU, W. (2023). Sampling without replacement from a high-dimensional finite population. *Bernoulli* **29**, 3198–3220.
[16] JEONG, M., DYTSO, A. AND CARDONE, M. (2021). Retrieving data permutations from noisy observations: High and low noise asymptotics. In *Proc. 2021 IEEE Int. Symp. Information Theory (ISIT)*, pp. 1100–1105.
[17] JOLLIFFE, I. T. (2002). *Principal Component Analysis*, 2nd ed,. Springer, New York.
[18] JONSSON, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12**, 1–38.
[19] LIANG, G., WILKES, D. AND CADZOW, J. (1993). ARMA model order estimation based on the eigenvalues of the covariance matrix. *IEEE Trans. Sig. Proc.* 41, 3003–3009.
[20] MA, R., CAI, T. T. AND LI, H. (2021). Optimal permutation recovery in permuted monotone matrix model. *J. Amer. Statist. Assoc.* **116**, 1358–1372.
[21] MARČENKO, V. AND PASTUR, L. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik* **114**, 507–536.
[22] SILVERSTEIN, J. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55**, 331–339.
[23] SILVERSTEIN, J. (2009). The Stieltjes transform and its role in eigenvalue behavior of large dimensional random matrices. In *Random Matrix Theory and its Applications* (Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singapore **18**). World Scientific, Hackensack, NJ, pp. 1–25.
[24] SILVERSTEIN, J. AND BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54**, 175–192.
[25] SILVERSTEIN, J. AND CHOI, S. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* **54**, 295–309.
[26] WACHTER, K. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Prob.* **6**, 1–18.
[27] WAX, M. AND KAILATH, T. (1985). Detection of signals by information theoretic criteria. *IEEE Trans. Acoust. Speech Sig. Proc.* **33**, 387–392.
[28] YAO, J., ZHENG, S. AND BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis* (Camb. Ser. Statist. Probabilistic Math. **39**). Cambridge University Press.
[29] YIN, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20**, 50–68.