Statistica Sinica **18**(2008), 425-442

LARGE SAMPLE COVARIANCE MATRICES WITHOUT INDEPENDENCE STRUCTURES IN COLUMNS

Zhidong Bai[1,2] and Wang Zhou[2]

[1] *Northeast Normal University and* [2] *National University of Singapore*

Abstract: The limiting spectral distribution of large sample covariance matrices is derived under dependence conditions. As applications, we obtain the limiting spectral distributions of Spearman's rank correlation matrices, sample correlation matrices, sample covariance matrices from finite populations, and sample covariance matrices from causal AR(1) models.

Key words and phrases: AR(1) model, finite population, limiting spectral distribution, random matrix theory, sample correlation matrices, sample covariance matrices, Spearman's rank correlation matrices.

1. Introduction

Let $\{X_{jk}^{(n)},\ j = 1,\ldots,p,\ k = 1,\ldots,n\}$, be an array of complex random variables for each n. Write $\mathcal{X}_n = (X_{jk})_{1\le j\le p, 1\le k\le n}$. Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be n columns of $\mathcal{X}_n$. Throughout the paper, we assume that the $\mathbf{X}_k' s$ are independent. Write

$$\mathbf{B}_n = \frac{1}{n}\mathcal{X}_n\mathcal{X}_n^* = \sum_{k=1}^{n}\mathbf{r}_k\mathbf{r}_k^*, \quad \mathbf{B}_{k,n} = \mathbf{B}_n - \mathbf{r}_k\mathbf{r}_k^*,$$

where $\mathbf{r}_k = n^{-1/2}\mathbf{X}_k$, $k = 1,\ldots,n$. $\mathbf{B}_n$ is the so-called *sample covariance* matrix. It should be noted that in the construction of the sample covariance matrix, the sample mean vector $\overline{\mathbf{X}} = n^{-1}\sum_{j=1}^{n}\mathbf{X}_j$ is not subtracted from $\mathbf{X}_k$ since it does not affect the limiting spectral distributions (LSD). The reasoning refers to the rank inequality (see Lemma 2.6 in Bai (1999)).

Sample covariance matrices are very important in multivariate statistical inference since many test statistics are defined by their eigenvalues or functionals. Under the assumption that all variables X_{jk} are independent and identically distributed (i.i.d.), the spectral analysis of large-dimensional sample covariance matrices has been actively developed since the pioneering work of Marčenko and Pastur (1967). Extensions can be found in the remarkable work of Wachter (1978) and Yin (1986). We also refer to the review paper of Bai (1999).

To relax the independence of the entries of $\mathbf{X}_k$, Silverstein (1995) considered the case of $\mathbf{X}_k = \mathbf{T}^{1/2}\mathbf{Y}_k$, where $\mathbf{T}$ is a non-negative definite matrix and $\mathbf{Y}_k$ consists of i.i.d. entries. Some further investigation on this model can be found in

Silverstein and Bai (1995) for strong convergence of the empirical spectral distribution (ESD) of sample covariance matrices, in Bai and Silverstein (1998, 1999) for spectrum separation, and in Bai and Silverstein (2004) for a central limit theorem for linear spectral analysis of sample covariance matrices. Naturally, one may ask whether we can completely get rid of the independence structure, since multivariate populations without independence structures can be found in many practical situations, e.g., Spearman's rank correlation matrix discussed later in the paper. The first attempt was made in Yin and Krishnaiah (1986), where the vectors $\mathbf{X}_k$ are distributed isotropically. Extensions to products of a non-negative definite matrix $\mathbf{T}$ with a sample covariance matrix when the underlying distribution is isotropic are given in Bai, Yin and Krishnaiah (1986). In this paper, we consider LSD of large sample covariance matrices under a very general dependence structure. As applications of our main theorem, we obtain the LSD of Spearman's rank correlation matrices, sample correlation matrices without second moment, sample covariance matrices from finite populations, and sample covariance matrices from causal AR(1) models. These models cannot be expressed as Silverstein's sample covariance matrices under independent structures.

Here is some notation. The eigenvalues of $\mathbf{B}_n$ are denoted by $\lambda_1, \ldots, \lambda_p$. The ESD of $\mathbf{B}_n$ is defined as

$$F^{\mathbf{B}_n}(x) = \frac{1}{p}\sum_{k=1}^{p} \mathbf{I}(\lambda_k \le x).$$

And the Stieltjes transform of $F^{\mathbf{B}_n}$ is given by

$$m_n(z) = m_{F^{\mathbf{B}_n}}(z) = \int \frac{1}{x-z} dF^{\mathbf{B}_n}(x) = \frac{1}{p}\mathrm{tr}(\mathbf{B}_n - z\mathbf{I})^{-1},$$

where $z = u + iv \in \mathcal{C}^+$, and $\mathbf{I}$ is the identity matrix.

In the following, we use $\|\cdot\|$ to denote the sup norm for bounded functions, the Euclidean norm for vectors, and the spectral norm for matrices. We also write $\mathcal{B}_n = \mathbf{B}_n - z\mathbf{I}$, $\mathcal{B}_{k,n} = \mathbf{B}_{k,n} - z\mathbf{I}$.

Theorem 1.1. *As $n \to \infty$, assume the following.*

1. *For all k, $\mathbf{E}\bar{X}_{jk}X_{lk} = t_{lj}$, and for any non-random $p \times p$ matrix $B = (b_{jk})$ with bounded norm, $\mathbf{E}|\mathbf{X}_k^* B\mathbf{X}_k - tr(B\mathbf{T})|^2 = o(n^2)$, where $\mathbf{T} = \mathbf{T}_n = (t_{jl})$.*
2. *$c_n := p/n \to c \in (0, \infty)$.*
3. *The norm of the matrix $\mathbf{T} = \mathbf{T}_n$ is uniformly bounded and $F^{\mathbf{T}}$ tends to a non-random probability distribution H.*

Then, with probability 1, $F^{\mathbf{B}_n}$ tends to a probability distribution, whose Stieltjes transform $m = m(z)$ $(z \in \mathcal{C}^+)$ satisfies

$$m = \int \frac{1}{t(1-c-czm) - z} dH(t). \tag{1.1}$$

If $\underline{m}(z) = -(1-c)/z + cm(z)$, then (1.1) becomes

$$z = -\frac{1}{\underline{m}} + c\int \frac{t}{1+\underline{m}t} dH(t), \tag{1.2}$$

which gives an explicit inverse function. Sometimes, it is easier to solve (1.2) than (1.1).

For the purpose of applications, we have the following.

Corollary 1.1. *If*

$$\mathbf{E}\bar{X}_{jk}X_{lk} = t_{lj}, \quad \textit{for all } \ k, \tag{1.3}$$

$$n^{-1}\max_{j\neq l}\mathbf{E}\Big|\bar{X}_{jk}X_{lk} - t_{lj}\Big|^2 \to 0 \textit{ uniformly in } \ k \le n, \tag{1.4}$$

$$n^{-2}\sum_{\Lambda}\Big(\mathbf{E}(\bar{X}_{jk}X_{lk} - t_{lj})(X_{j'k}\bar{X}_{l'k} - t_{j'l'})\Big)^2 \to 0 \textit{ uniformly in } \ k \le n, \tag{1.5}$$

where $\Lambda = \{(j,l,j',l') : 1 \le j,l,j',l' \le p\} \setminus \{(j,l,j',l') : j = j' \ne l = l'$ *or* $j = l' \ne j' = l\}$, *then* 1 *of Theorem* 1.1 *holds. Consequently, Theorem* 1.1 *is true if we replace these moment conditions by* (1.3), (1.4) *and* (1.5).

Remark 1.1. When $\mathbf{T}_n = \mathbf{I}_p$, the identity matrix, $F^{\mathbf{B}_n}$ tends to the Marčenko-Pastur (MP) law, whose Stieltjes transform is

$$m(z) = \frac{1-c-z+\sqrt{(1+c-z)^2-4c}}{2cz} \tag{1.6}$$

where, in accordance with Bai (1993), the square root of a complex number is defined to be the one with a positive imaginary part. Equation (1.1) then reduces to the quadratic equation $czm^2(z) + (c+z-1)m(z) + 1 = 0$.

In Theorem 1.1, it does not matter if $\mathbf{E}X_{jk} = 0$ or not, because Lemma 2.6 in Bai (1999) implies that, for $A = n^{-1/2}\mathcal{X}_n$ and $B = A - \mathbf{E}A$,

$$\|F^{AA^*} - F^{BB^*}\| \le \frac{1}{n}\text{rank}(A-B) = \frac{1}{n}.$$

The proofs of Theorem 1.1 and Corollary 1.1 are deferred to Section 3.

2. Applications

2.1. Spearman's rank correlation matrices

Suppose (X,Y), (X_j,Y_j), $j = 1,\ldots,n$ are i.i.d. random vectors from a bivariate normal distribution $F(x,y)$. Define the correlation coefficient by

$$\rho = \frac{\mathbf{E}(X-\mathbf{E}X)(Y-\mathbf{E}Y)}{\sqrt{\mathbf{E}(X-\mathbf{E}X)^2(Y-\mathbf{E}Y)^2}}.$$

We wish to test for independence, or to test $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$. One test statistic is the sample correlation coefficient

$$\rho_{XY} = \frac{\sum_{j=1}^n (X_j - \overline{X})(Y_j - \overline{Y})}{\sqrt{\sum_{j=1}^n (X_j - \overline{X})^2 \sum_{j=1}^n (Y_j - \overline{Y})^2}}, \tag{2.1}$$

where $\overline{X} = \sum_{j=1}^n X_j/n, \overline{Y} = \sum_{i=j}^n Y_j/n$. If strong structural assumptions on the population, such as normality, are not present, some nonparametric measures of independence must be proposed. One of them is based on the Spearman's rank correlation coefficient.

To calculate the Spearman's rank correlation coefficient r_s, let Q_j denote the rank of X_j among $\{X_1, \ldots, X_n\}$ and let S_j be the rank of Y_j among $\{Y_1, \ldots, Y_n\}$. Replace X_j's and Y_j's in (2.1) by Q_j's and S_j's to get *Spearman's rank correlation coefficient* (Spearman (1904))

$$r_s = \frac{\sum Q_j S_j - \frac{1}{n}(\sum Q_j)(\sum S_j)}{\sqrt{\sum (Q_j - \overline{Q})^2 \sum (S_j - \overline{S})^2}} = \frac{12}{n(n^2-1)} \sum (Q_j - \overline{Q})(S_j - \overline{S}),$$

where $\overline{Q} = \overline{S} = (n+1)/2$. The most important point is that r_s is distribution free. For more statistical properties of r_s, the reader is referred to Hájek, Šidák and Sen (1999).

Now let us generalize the two-dimensional problem to the large-dimensional case. Suppose that we have n i.i.d. samples $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, where $\mathbf{Y}_k$ is a p-vector consisting of i.i.d. random variables $Y_{1k}, \ldots, Y_{pk}$, $k = 1, \ldots, n$. The question is how to test the independence among the components of $\mathbf{Y}_k's$. Consider Spearman's rank correlation matrices, $\mathcal{R}_p = (r_{kl})_{1 \le k,l \le p}$, where r_{kl} is Spearman's rank correlation coefficient between the k-th and l-th rows of $[\mathbf{Y}_1, \ldots, \mathbf{Y}_n]$. Here, as an application of our Corollary 1.1, we derive the limiting spectral distribution of $\mathcal{R}_p$, when $c_n = p/n \to c$.

Actually, $\mathcal{R}_p$ can be expressed in the form $p^{-1}\mathcal{X}_p^T \mathcal{X}_p$. For that purpose, we denote by Q_{jk} the rank of Y_{jk} among $\{Y_{j1}, \ldots, Y_{jn}\}$. For $j = 1, \ldots, p$, write

$$\mathbf{X}_j = \frac{\sqrt{12p}}{\sqrt{n(n^2-1)}} \Big(Q_{j1} - \frac{n+1}{2}, \ldots, Q_{jn} - \frac{n+1}{2}\Big)^T,$$

and $\mathcal{X}_p = (\mathbf{X}_1, \ldots, \mathbf{X}_p)$. Then $\mathcal{R}_p = p^{-1}\mathcal{X}_p^T \mathcal{X}_p$, by noticing that $\mathcal{R}_p = (p^{-1}\mathbf{X}_i^T \mathbf{X}_j)_{p \times p}$. Since the spectra of $\mathcal{X}_p^T \mathcal{X}_p$ and $\mathcal{X}_p \mathcal{X}_p^T$ differ by $|n-p|$ zero eigenvalues, it suffices to consider $F^{p^{-1}\mathcal{X}_p \mathcal{X}_p^T}$.

By elementary calculation, we have

$$t_{jl} = \mathbf{E} X_{jk} X_{lk} = \begin{cases} pn^{-1}, & \text{if } j = l \\ -pn^{-1}(n-1)^{-1}, & \text{otherwise.} \end{cases}$$

From this, one can see that $\mathbf{T}$ has one eigenvalue of 0 and $n-1$ of $p/(n-1)$, which verifies that the norm of $\mathbf{T}$ is uniformly bounded and H is a degenerate distribution with mass at $\{c\}$. To apply Corollary 1.1, we need to verify

$$\sum_{\Lambda}\Big(b\mathbf{E}(X_{jk}X_{lk}-t_{jl})(X_{j'k}X_{l'k}-t_{j'l'})\Big)^2=o(n^2), \tag{2.2}$$

$$n^{-1}\max_{j\neq l}\mathbf{E}|X_{jk}X_{lk}-t_{jl}|^2\to 0, \tag{2.3}$$

where $\Lambda=\{(j,l,j',l'):1\le j,l,j',l'\le n\}\setminus\{(j,l,j',l'):j=j'\neq l=l'$ or $j=l'\neq j'=l\}$.

After some elementary but tedious calculations, we have

$$\begin{aligned}&\mathbf{E}X_{11}^4\sim\frac{18p^2}{5n^2},\quad \mathbf{E}X_{11}^2=\frac{p}{n},\\&\mathbf{E}X_{11}^2X_{21}X_{31}\sim-\frac{p^2}{n^3},\quad \mathbf{E}X_{11}X_{21}X_{31}X_{41}\sim\frac{3p^2}{n^4},\\&\mathbf{E}(X_{11}^2-\mathbf{E}X_{11}^2)(X_{21}^2-\mathbf{E}X_{21}^2)=O(n^{-1}),\end{aligned}$$

as $n\to\infty$, where $a\sim b$ means $a/b\to 1$. (2.2)−(2.3) are direct consequences of the above relations.

Since $n/p\to 1/c$, Corollary 1.1 implies $F^{p^{-1}\mathcal{X}_p\mathcal{X}_p^T}$ tends almost surely to a limiting law F whose Stieltjes transform is given by

$$m=\frac{1}{c(1-(1/c)-(1/c)zm)-z}=\frac{1}{c-1-zm-z}.$$

From this, we have

$$m(z)=\frac{-(1-c+z)+\sqrt{(1+c-z)^2-4c}}{2z}.$$

Noting that

$$F^{p^{-1}\mathcal{X}_p\mathcal{X}_p^T}=(1-\frac{p}{n})I[0,\infty)+\frac{p}{n}F^{p^{-1}\mathcal{X}_p^T\mathcal{X}_p}, \tag{2.4}$$

the limiting law $\underline{F}$ of $F^{p^{-1}\mathcal{X}_p^T\mathcal{X}_p}$ should satisfy $F=(1-c)I[0,\infty)+c\underline{F}$ and its Stieltjes transform should satisfy $m(z)=-(1-c)z^{-1}+c\underline{m}(z)$.

Thus, $\underline{m}(z)=(1-c-z+\sqrt{(1+c-z)^2-4c})/(2cz)$ and we have proved the following theorem for Spearman's rank correlation matrix.

Theorem 2.2. *Suppose that all the Y_{ik}, $i=1,\ldots,p$, $k=1,\ldots,n$ are independent and have a continuous distribution. Then $F^{\mathcal{R}_p}$ tends to the MP law a.s. as $p/n\to c$, with Stieltjes transform given by*

$$m_{\mathcal{R}_n}(z)=\frac{1-c-z+\sqrt{(1+c-z)^2-4c}}{2cz}. \tag{2.5}$$

Remark 2.2. We assume the continuity of the random variables in order to avoid ties among the observations. Otherwise, the number of ties must be taken into account. As a consequence, the expectation and the variance under the null (independence) hypothesis need adjustment. We refer to Hollander and Wolfe (1999) for more details. For the adjusted r_{ij}, Theorem 2.2 still holds.

2.2. Sample correlation matrices

Spectral distributions of sample correlation matrices were first studied by Jiang (2004) under second moment conditions. In this part, we derive the MP law for sample correlation matrices under conditions weaker than the second moment.

Suppose $\{Y, Y_{jk}, j,\ k = 1, \ldots\}$ are i.i.d random variables with mean 0. Write $\mathcal{Y}_n = (Y_{jk})_{1\le j\le p, 1\le k\le n}$. Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be n columns of $\mathcal{Y}_n$. Then $\mathcal{Y}_n = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$. From the statistical point of view, $\mathbf{Y}_k$ consists of p observations of the kth component of the multivariate population. Hence the n by n sample correlation matrix is $\mathcal{R}_Y = n^{-1}\mathcal{X}_n^T\mathcal{X}_n$ with $\mathcal{X}_n = n^{1/2}(\mathbf{Y}_1/\|\mathbf{Y}_1\|, \ldots, \mathbf{Y}_n/\|\mathbf{Y}_n\|)$, where $\|\cdot\|$ is the usual Euclidean norm.

Theorem 2.3. *Assume that Y belongs to the attraction domain of the normal law. Then $F^{\mathcal{R}_Y}$ tends to the MP law a.s. as $p/n \to c$, with Stieltjes transform*

$$m_{\mathcal{R}_Y} = \frac{-(cz - c + 1) + \sqrt{(cz - c - 1)^2 - 4c}}{2z}. \tag{2.6}$$

As in the derivation of Theorem 2.2, we still consider $F^{n^{-1}\mathcal{X}_n\mathcal{X}_n^T}$. In order to apply Corollary 1.1, let us check moment conditions for $\mathcal{X}_n$ first. Write $X_{jk} = n^{1/2}Y_{jk}/\|\mathbf{Y}_k\|$. From (3.7) and (3.10) of Giné, Götze and Mason (1997), we have

$$n^{-2}\mathbf{E}X_{11}^4 = o(p^{-1}), \quad n^{-2}\mathbf{E}X_{11}^2X_{21}X_{31} = o(p^{-3}), \tag{2.7}$$

$$n^{-2}\mathbf{E}X_{11}X_{21}X_{31}X_{41} = o(p^{-4}). \tag{2.8}$$

Since

$$1 = n^{-2}\Big(\sum_{i=1}^p X_{j1}^2\Big)^2 = n^{-2}\sum_{j=1}^p X_{j1}^4 + n^{-2}\sum_{1\le j\ne l\le n} X_{j1}^2X_{j1}^2,$$

we have $1 = n^{-2}p\mathbf{E}X_{11}^4 + n^{-2}p(p-1)\mathbf{E}X_{11}^2X_{21}^2$ which, combined with $n^{-2}\mathbf{E}X_{11}^4 = o(p^{-1})$, implies $n^{-2}\mathbf{E}X_{11}^2X_{21}^2 = O(p^{-2})$ as $n \to \infty$. From (2.7) and $n^{-1}t_{jl} = o(p^{-2})$ for $j \ne l$ (see (3.12) of Giné, Götze and Mason (1997)), we have, uniformly in $k \le n$ as $n \to \infty$,

$$\sum_{\Lambda} |\mathbf{E}(X_{jk}X_{lk} - t_{jl})(X_{j'k}X_{l'k} - t_{j'l'})|^2 = o(n^2),$$

$$n^{-1} \max_{j \neq l} \mathbf{E}|X_{jk}X_{lk} - t_{jl}|^2 \to 0,$$

where $\Lambda = \{(j,l,j',l') : 1 \leq j,l,j',l' \leq p\} \setminus \{(j,l,j',l') : j = j' \neq l = l'$ or $j = l' \neq j' = l\}$.

Since $F^{\mathbf{T}}(x) = p^{-1}\delta_{t_{11}+(p-1)t_{12}}(x) + (p-1)p^{-1}\delta_{t_{11}-t_{12}}(x)$, the norm of $\mathbf{T}$ is bounded and H is the degenerate distribution with whole mass at $1/c$. So all the assumptions of Corollary 1.1 hold for $n^{-1}\mathcal{X}_n\mathcal{X}_n^T$. Hence the Stieltjes transform of the limit of $F^{n^{-1}\mathcal{X}_n\mathcal{X}_n^T}$ is

$$m = \frac{1}{c^{-1}(1-c-czm) - z} = \frac{c}{1-c-cz-czm},$$

from which is followed that

$$m = \frac{-(cz+c-1) + \sqrt{(cz-c-1)^2 - 4c}}{2cz}.$$

Since $F^{n^{-1}\mathcal{X}_n^T\mathcal{X}_n} = (1-p/n)I[0,\infty) + (p/n)F^{n^{-1}\mathcal{X}_n\mathcal{X}_n^T}$, we have $m_{\mathcal{R}_Y} = [-(cz - c + 1) + \sqrt{(cz-c-1)^2 - 4c}]/(2z)$.

2.3. Sample covariance matrix for a finite population

Suppose $\mathbf{y} = (y_1, \ldots, y_p)^T$ is a simple random sample of size n from a finite population of size N with values $\{u_1, \ldots, u_N\}$. Let $\mathbf{y}_1 = (y_{11}, \ldots, y_{p1})^T, \ldots, \mathbf{y}_n = (y_{1n}, \ldots, y_{pn})^T$ be n independent copies of $\mathbf{y}$. Without loss of generality, we assume $\mathbf{E}y_j = \sum_{s=1}^N u_s/N = 0$. Write $\sigma^2 = \sum_{s=1}^N u_s^2/N$, $X_i = y_i/\sigma$, $i = 1, \ldots, n$. Then after some algebra, we have $\mathbf{E}X_1^4 = \mathbf{E}y_1^4/\sigma^4$, $\mathbf{E}X_1^2 = 1$, and

$$\begin{aligned}\mathbf{E}X_1^2X_2X_3 &= \frac{1}{\sigma^4}\frac{1}{N(N-1)(N-2)}\sum_{1\leq s\neq t\neq l\leq N} u_s^2u_tu_l \\ &= \frac{1}{\sigma^4}\left(\frac{2}{(N-1)(N-2)}\mathbf{E}y_1^4 - \frac{N}{(N-1)(N-2)}\sigma^4\right),\end{aligned}$$

$$\begin{aligned}\mathbf{E}X_1X_2X_3X_4 &= \frac{1}{\sigma^4}\frac{1}{N(N-1)(N-2)(N-3)}\sum_{1\leq s\neq t\neq l\neq m\leq N} u_su_tu_lu_m \\ &= \frac{n}{(N-1)(N-2)(N-3)\sigma^4}(-12\mathbf{E}y_1^4 + 3N\sigma^4),\end{aligned}$$

$$\begin{aligned}\mathbf{E}(X_1^2 - \mathbf{E}X_1^2)(X_2^2 - \mathbf{E}X_2^2) &= \frac{1}{\sigma^4}\left(\frac{1}{N(N-1)}\sum_{1\leq s\neq t\leq N} u_s^2u_t^2 - \sigma^4\right) \\ &= \frac{1}{\sigma^4}\frac{1}{N-1}(\sigma^4 - \mathbf{E}y_1^4).\end{aligned}$$

Since $n \le N$, all the moment assumptions in Corollary 1.1 are satisfied if we suppose $\limsup_{N\to\infty} \mathbf{E}y_1^4 < \infty$ and $\liminf_{N\to\infty} \sigma^2 > 0$.

Simple calculations show that $t_{jl} = -1/(N-1)$ for $j \ne l$. Hence one eigenvalue of $\mathbf{T}$ is $(N-p)(N-1)^{-1}$ and the other $p-1$ are $N(N-1)^{-1}$, which implies that the norm of $\mathbf{T}$ is bounded and H is a distribution concentrated on $\{1\}$.

Now we see that all the conditions of Corollary 1.1 are satisfied. Hence the Stieltjes transform of the limit of $F^{n^{-1}\mathcal{X}_n\mathcal{X}_n^T}$ is given by

$$m = \frac{1}{(1-c-czm)-z},$$

from which is followed that

$$m = \frac{-(z+c-1)+\sqrt{(z-c-1)^2-4c}}{2cz}.$$

Since $F^{n^{-1}\mathcal{X}_n^T\mathcal{X}_n} = (1-p/n)I[0,\infty) + (p/n)F^{n^{-1}\mathcal{X}_n\mathcal{X}_n^T}$, we have

$$m_{n^{-1}\mathcal{X}_n^T\mathcal{X}_n} = \frac{-(z-c+1)+\sqrt{(z-c-1)^2-4c}}{2z}. \tag{2.9}$$

Theorem 2.4. *Assume that* $\limsup_{N\to\infty} \mathbf{E}y_1^4 < \infty$ *and* $\liminf_{N\to\infty} \sigma^2 > 0$. *If* $p/n \to c$, *then* $F^{n^{-1}\mathcal{X}_n^T\mathcal{X}_n}$ *tends to the MP law a.s., with Stieltjes transform given by* (2.9).

2.4. Sample covariance matrix from a causal time series model

We first consider the simple $AR(1)$ model. Suppose that $\mathbf{y} = (y_1,\ldots,y_p)^T$ is a sample of size p from a causal $AR(1)$ model, i.e.,

$$y_t = \phi y_{t-1} + \varepsilon_t$$

where $\phi \in (-1,1)$ is a constant and $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean 0 and variance 1. Let $\mathbf{X}_1 = (X_{11},\ldots,X_{p1})^T,\ldots,\mathbf{X}_n = (X_{1n},\ldots, X_{pn})^T$ be n independent copies of $\mathbf{y}$.

Theorem 2.5. *Assume that the innovations* ε_j *have mean* 0, *variance* 1 *and finite* 4*th moment. Then, the LSD of* $\mathbf{B}_n$ *exists and its Stieltjes transform is given by* $m = \underline{m}/c + (1-c)/(cz)$, *where* $\underline{m}$ *is the unique solution in the upper complex plane to the* 4*th degree polynomial equation*

$$(z\underline{m}+1)^2(\underline{m}^2 + 2\underline{m}(1+\phi^2) + (1-\phi^2)^2) = c^2\underline{m}^2. \tag{2.10}$$

It is well known that

$$\mathbf{T} = \frac{1}{1-\phi^2}\begin{pmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{p-1} \\ \phi & 1 & \phi & \cdots & \phi^{p-2} \\ \phi^2 & \phi & 1 & \cdots & \phi^{p-3} \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \phi^{p-1} & \phi^{p-2} & \phi^{p-3} & \cdots & 1 \end{pmatrix}.$$

By the Gerschgorn Theorem, the largest eigenvalue of $\mathbf{T}$ is not larger than $(1-\phi)^{-2}[1+2(|\phi|+\phi^2+\cdots)] \le 2(1-\phi^2)^{-2}$.

Next, we verify the moment condition of Theorem 1.1. By the causal expression, for $j \ge l$, we have

$$X_{jk}X_{lk} - t_{jl} = \sum_{h=0}^{\infty} \phi^{j-l+2h}(\varepsilon_{l-h}^2 - 1) + \sum_{\substack{j-l\neq h_1-h_2 \\ h_1,h_2\ge 0}} \varepsilon_{j-h_1}\varepsilon_{l-h_2}\phi^{h_1+h_2}. \tag{2.11}$$

Let $j \ge l$, $j' \ge l'$ and $l \ge l'$. By (2.11), we have

$$\mathbf{E}(X_{jk}X_{lk} - t_{jl})(X_{j'k}X_{l'k} - t_{j'l'}) = \frac{(E\varepsilon_1^4 - 3)\phi^{j+j'+l-3l'}}{1-\phi^4} + t_{jj'}t_{ll'} + t_{jl'}t_{j'l}.$$

If $j \ge l$, $j' \ge l'$ and $l' \ge l$. We can similarly show that

$$\mathbf{E}(X_{jk}X_{lk} - t_{jl})(X_{j'k}X_{l'k} - t_{j'l'}) = \frac{(E\varepsilon_1^4 - 3)\phi^{j+j'+l'-3l}}{1-\phi^4} + t_{jj'}t_{ll'} + t_{jl'}t_{j'l}.$$

Using the same argument in the other six cases, one can show that

$$\mathbf{E}(X_{jk}X_{lk} - t_{jl})(X_{j'k}X_{l'k} - t_{j'l'}) = \frac{(E\varepsilon_1^4 - 3)\phi^{\mu_1+\mu_2+\mu_4-3\mu_4}}{1-\phi^4} + t_{jj'}t_{ll'} + t_{jl'}t_{j'l}.$$

where, for $t = 1, 2, 3, 4$, μ_t is the t-th largest value among $\{j, j', l, l'\}$. Let $\mathbf{B}$ be a non-random matrix with bounded norm. Since the entries of $\mathbf{B}$ are bounded, we have

$$\sum_{j,j',l,l'} b_{jl}b_{j',l'}\frac{(E\varepsilon_1^4 - 3)\phi^{\mu_1+\mu_2+\mu_4-3\mu_4}}{1-\phi^3} = O(n),$$

$$\sum_{j,j',l,l'} b_{jl}b_{j',l'}[t_{jj'}t_{ll'} + t_{jl'}t_{j'l}] = tr\mathbf{TBTB} + tr\mathbf{TBTB}^T = O(n).$$

Thus, the moment condition of Theorem 1.1 is satisfied.

To apply the main theorem, we need to find the LSD of $\mathbf{T}$. To this end, we use

$$\mathbf{T}^{-1} = \begin{pmatrix} 1 & -\phi & 0 & \cdots & 0 & 0 \\ -\phi & 1+\phi^2 & -\rho & \cdots & 0 & \\ 0 & -\phi & 1+\phi^2 & \cdots & 0 & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots & \\ 0 & 0 & 0 & \cdots & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{pmatrix}.$$

Make a slight modification to $\mathbf{T}^{-1}$ as

$$\widehat{\mathbf{T}}^{-1} = \begin{pmatrix} 1+\phi^2 & -\phi & 0 & \cdots & 0 & 0 \\ -\phi & 1+\phi^2 & -\rho & \cdots & 0 & \\ 0 & -\phi & 1+\phi^2 & \cdots & 0 & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots & \\ 0 & 0 & 0 & \cdots & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & \cdots & -\phi & 1+\phi^2 \end{pmatrix}.$$

By Lemma 2.2 of Bai (1999), the LSD of $\mathbf{T}^{-1}$ is the same as that $\widehat{\mathbf{T}}^{-1}$. It is not difficult to show that the eigenvalues of $\widehat{\mathbf{T}}^{-1}$ are

$$1+\phi^2+2\phi\cos(k\pi/(p+1)), \quad k=1,\ldots,p. \tag{2.12}$$

Then, the equation (1.6) for this case is

$$\begin{aligned} z &= -\frac{1}{\underline{m}} + c\int_0^1 \frac{dt}{\underline{m}+1+\phi^2+2\phi\cos(\pi t)} \\ &= -\frac{1}{\underline{m}} + \frac{c}{2\pi i}\oint_{|\zeta|=1} \frac{d\zeta}{\zeta[\underline{m}+1+\phi^2+\phi(\zeta+\zeta^{-1})]} \\ &= -\frac{1}{\underline{m}} - \frac{c}{\sqrt{(\underline{m}+1+\phi^2)^2-4\phi^2}}. \end{aligned}$$

This can be simplified to (2.10). The proof is complete.

All results about the $AR(1)$ model, except for the explicit expression of the equation for the Stieltjes transform, can be easily extended to the causal time series model. Let

$$y_t = \sum_{h=0}^{\infty} \psi(h)\varepsilon_{t-h}$$

with $\sum_h |\psi(h)| = L < \infty$. Then $T = (t_{ij})_{n\times n}$ with $t_{ij} = \sum_{h=0}^{\infty} \psi(h)\psi(|i-j|+h)$. By the Gerschgorn Theorem, one can show that the norm of the matrix T is not larger than $2L^2$. It is easy to show that for each fixed k, $n^{-1}tr(T^k) \to \nu_k \leq 2^k L^{2k}$, which shows the Carleman condition is satisfied. By the Moment Convergence Theorem, the ESD of T tends to a limit H.

Similar to the $AR(1)$ case, one can show that

$$\begin{aligned}&\mathbf{E}(X_{jk}X_{lk} - t_{jl})(X_{j'k}X_{l'k} - t_{j'l'})\\&\quad = (E\varepsilon_1^4 - 3)\sum_{h=0}^{\infty}\psi(\mu_1 - \mu_4 + h)\psi(\mu_2 - \mu_4 + h)\psi(\mu_3 - \mu_4 + h)\psi(h)\\&\qquad + t_{jj'}t_{ll'} + t_{jl'}t_{j'l}.\end{aligned}$$

From this, one can check the moment condition of Theorem 1.1. Therefore, the LSD of $\mathbf{B}_n$ exists and its Stieltjes transform is given by (1.1).

3. Proof of Theorem 1.1 and Corollary 1.1

Proof of Theorem 1.1. We now proceed with the proof by the following steps.
1. $m_n(z) - \mathbf{E}m_n(z) \to 0$, a.s..
2. $\mathbf{E}m_n(z) \to m$, which satisfies (1.1).
3. The equation has a unique solution in $\mathcal{C}^+$.

Step 1. Proof of $m_n(z) - \mathbf{E}m_n(z) \to 0$, a.s..

$\mathbf{E}_k$ denotes the conditional expectation given $\mathbf{X}_{k+1}, \ldots, \mathbf{X}_n$. With this notation, we have $m_n(z) = \mathbf{E}_0(m_n(z))$, $\mathbf{E}m_n(z) = \mathbf{E}_n(m_n(z))$. Therefore,

$$\begin{aligned}m_n(z) - \mathbf{E}m_n(z) &= \sum_{k=1}^{n}(\mathbf{E}_{k-1}(m_n(z)) - \mathbf{E}_k(m_n(z)))\\&= \frac{1}{p}\sum_{k=1}^{n}[\mathbf{E}_{k-1} - \mathbf{E}_k]\Big(\mathrm{tr}\mathcal{B}_n^{-1} - \mathrm{tr}\mathcal{B}_{k,n}^{-1}\Big)\\&= \frac{1}{p}\sum_{k=1}^{n}[\mathbf{E}_{k-1} - \mathbf{E}_k]\gamma_k,\end{aligned}$$

where

$$\gamma_k = \frac{\mathbf{r}_k^*\mathcal{B}_{k,n}^{-2}\mathbf{r}_k}{1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k}.$$

Since $|\gamma_k| \leq v^{-1}$, $\{[\mathbf{E}_{k-1} - \mathbf{E}_k]\gamma_k\}$ forms a bounded martingale difference sequence. Applying Burkholder inequality, we have

$$\mathbf{E}|m_n(z) - \mathbf{E}m_n(z)|^q \leq K_q p^{-q}\mathbf{E}\bigg(\sum_{k=1}^{n}|(\mathbf{E}_{k-1} - \mathbf{E}_k)\gamma_k|^2\bigg)^{\frac{q}{2}}$$

$$\leq K_q(\frac{2}{v})^q p^{-\frac{q}{2}}(\frac{p}{n})^{-\frac{q}{2}} \tag{3.1}$$

which, for $q > 2$, implies $p^{-1}\sum_{k=1}^{n}(\mathbf{E}_{k-1} - \mathbf{E}_k)\gamma_k \to 0$, a.s..

Step 2. Proof of $\mathrm{E}m_n(z) \to m$, which satisfies (1.1).

Write $\mathbf{K} = \mathbf{T}(1 + n^{-1}\mathrm{tr}\mathcal{B}_{k,n}^{-1}\mathbf{T})^{-1}$. Since

$$(\mathbf{B}_n - z\mathbf{I}) - (\mathbf{K} - z\mathbf{I}) = \sum_{k=1}^{p}\mathbf{r}_k\mathbf{r}_k^* - \mathbf{K},$$

we have

$$\begin{aligned}
&(\mathbf{K} - z\mathbf{I})^{-1} - (\mathbf{B}_n - z\mathbf{I})^{-1} \\
&\quad= \sum_{k=1}^{p}(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{r}_k\mathbf{r}_k^*(\mathbf{B}_n - z\mathbf{I})^{-1} - (\mathbf{K} - z\mathbf{I})^{-1}\mathbf{K}(\mathbf{B}_n - z\mathbf{I})^{-1} \\
&\quad= \sum_{k=1}^{p}\frac{(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{r}_k\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}}{1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k} - (\mathbf{K} - z\mathbf{I})^{-1}\mathbf{K}(\mathbf{B}_n - z\mathbf{I})^{-1} \qquad (3.2)
\end{aligned}$$

where, in the last equation we have used the formula,

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

Multiplying $\mathbf{T}^\ell$ for $\ell = 0, 1$ on both sides of (3.2), we obtain

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

Taking the trace and dividing by p we find

$$\begin{aligned}
&\frac{1}{p}\mathrm{tr}\mathbf{T}^\ell(\mathbf{K} - z\mathbf{I})^{-1} - \frac{1}{p}\mathrm{tr}\Big(\mathbf{T}^\ell\mathcal{B}_n^{-1}\Big) \\
&\quad= \frac{1}{p}\sum_{k=1}^{n}\frac{\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{T}^\ell(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{r}_k}{1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k} - \frac{1}{p}\mathrm{tr}\mathbf{T}^\ell(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{K}\mathcal{B}_n^{-1} \\
&\quad= \frac{1}{p}\sum_{k=1}^{n}\frac{d_k}{1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k}, \qquad (3.3)
\end{aligned}$$

where

$$d_k = \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{T}^\ell(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{r}_k - \frac{1}{n}\mathrm{tr}\mathbf{T}^\ell(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{K}\mathcal{B}_n^{-1}(1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k).$$

Write $d_k = d_{k1} + d_{k2} + d_{k3}$ with

$$d_{k1} = \frac{1}{n}\mathrm{tr}(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{T}^{\ell+1}\mathcal{B}_{k,n}^{-1} - \frac{1}{n}\mathrm{tr}(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{T}^{\ell+1}\mathcal{B}_{n}^{-1},$$
$$d_{k2} = \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{T}^{\ell}(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{r}_k - \frac{1}{n}\mathrm{tr}\mathcal{B}_{k,n}^{-1}\mathbf{T}^{\ell+1}(\mathbf{K} - z\mathbf{I})^{-1},$$
$$d_{k3} = \frac{1}{n}\mathrm{tr}\Big((\mathbf{K} - z\mathbf{I})^{-1}\mathbf{T}^{\ell+1}\mathcal{B}_{n}^{-1}\Big(\frac{1 + \mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k}{1 + \frac{1}{n}\mathrm{tr}\mathcal{B}_{k,n}^{-1}\mathbf{T}} - 1\Big)\Big),$$

noting $\mathbf{T}(\mathbf{K} - z\mathbf{I})^{-1} = (\mathbf{K} - z\mathbf{I})^{-1}\mathbf{T}$.

We first show that

$$\|(\mathbf{K} - z\mathbf{I})^{-1}\| \le \frac{L}{v^2}, \tag{3.4}$$

for some constant L.

For any real t, we have

$$\Im\Big(t + z\Big(1 + \frac{1}{n}\mathrm{tr}\mathcal{B}_{k,n}^{-1}\mathbf{T}\Big)\Big) = v + \frac{v}{n}\mathrm{tr}\Big(\mathbf{B}_{k,n}\Big((\mathbf{B}_{k,n} - u\mathbf{I})^2 + v^2\Big)^{-1}\mathbf{T}\Big) > v,$$

noting that $\mathrm{tr}(\mathbf{B}_{k,n}(\mathbf{B}_{k,n} - u\mathbf{I})^2 + v^2)^{-1}\mathbf{T})$ is the trace of a non-negative definite Hermitian matrix since

$$\mathrm{tr}\Big(\mathbf{B}_{k,n}\Big((\mathbf{B}_{k,n} - u\mathbf{I})^2 + v^2\Big)^{-1}\mathbf{T}\Big) = \mathrm{tr}\Big(\mathbf{E}^k\mathbf{X}_k^*\mathbf{B}_{k,n}\Big((\mathbf{B}_{k,n} - u\mathbf{I})^2 + v^2\Big)^{-1}\mathbf{X}_k\Big),$$

where $\mathbf{E}^k$ denotes the conditional expectation given $\mathbf{X}_1, \ldots, \mathbf{X}_{k-1}, \mathbf{X}_{k+1}, \ldots, \mathbf{X}_n$.

Thus, using the spectral decomposition of $\mathbf{K}$,

$$\begin{aligned}
\|(\mathbf{K} - z\mathbf{I})^{-1}\| &\le \max_{t\ge 0}\left|\frac{t}{1 + \frac{1}{n}\mathrm{tr}\Big(\mathcal{B}_{k,n}^{-1}\mathbf{T}\Big)} - z\right|^{-1} \\
&\le \max_{t\ge 0}\left|\frac{1 + \frac{1}{n}\mathrm{tr}\Big(\mathcal{B}_{k,n}^{-1}\mathbf{T}\Big)}{t - z\Big(1 + \frac{1}{n}\mathrm{tr}\Big(\mathcal{B}_{k,n}^{-1}\mathbf{T}\Big)\Big)}\right|^{-1} \\
&\le \frac{L}{v^2},
\end{aligned}$$

where L can be chosen as any number $> v + \|\mathbf{T}\|$. The assertion (3.4) is proved.

Lemma 2.6 of Silverstein and Bai (1995) implies

$$|d_{k1}| \le \frac{\|(\mathbf{K} - z\mathbf{I})^{-1}\mathbf{T}^{\ell+1}\|}{nv} \le \frac{L}{nv^3} \to 0, \tag{3.5}$$

where L is a constant which may take different values in different appearances.

Next, we analyze d_{k2}. Write $\mathcal{B}_{k,n}^{-1}\mathbf{T}^{\ell}(\mathbf{K}-z\mathbf{I})^{-1}=(b_{jl})$. Then, by the moment conditions for $\mathcal{X}_n$, we have

$$\mathbf{E}|d_{k2}|^2=n^{-2}\mathbf{E}\Big|\sum_{j,l}b_{j,l}(\bar{X}_{jk}X_{lk}-t_{lj})\Big|^2\to 0. \tag{3.6}$$

Note that

$$|d_{k3}|\le\|(\mathbf{K}-z\mathbf{I})^{-1}\|\|\mathbf{T}\|^{\ell+1}|z|v^{-2}|\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k-\frac{1}{n}\mathrm{tr}\mathcal{B}_{k,n}^{-1}\mathbf{T}|.$$

Therefore, by moment conditions, we have

$$\mathbf{E}|d_{k3}|^2\le L\mathbf{E}|\mathbf{r}_k^*\mathcal{B}_{k,n}^{-1}\mathbf{r}_k-\frac{1}{n}\mathrm{tr}\mathcal{B}_{k,n}^{-1}\mathbf{T}|^2\to 0. \tag{3.7}$$

Notice that

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

Hence it follows from (3.3) and (3.5), (3.6), (3.7) that

$$\frac{1}{p}\Big(\mathbf{E}\mathrm{tr}\mathbf{T}^{\ell}(\mathbf{K}-z\mathbf{I})^{-1}-\mathbf{E}\mathrm{tr}\mathbf{T}^{\ell}\mathcal{B}_n^{-1}\Big)\to 0, \tag{3.8}$$

as $n\to\infty$.

Let $\widetilde{\mathbf{K}}=\mathbf{T}(1+n^{-1}\mathbf{E}\mathrm{tr}\mathbf{T}\mathcal{B}_n^{-1})^{-1}$. Similarly, we can prove that (3.4) holds also when $\mathbf{K}$ is replaced by $\widetilde{\mathbf{K}}$. Then

$$\begin{aligned}
&\frac{1}{p}\Big|\mathbf{E}\Big(\mathrm{tr}\mathbf{T}^{\ell}(\mathbf{K}-z\mathbf{I})^{-1}-\mathrm{tr}\mathbf{T}^{\ell}(\widetilde{\mathbf{K}}-z\mathbf{I})^{-1}\Big)\Big|\\
&\quad\le L^2\|\mathbf{T}\|^{\ell}v^{-4}\mathbf{E}\|\mathbf{K}-\widetilde{\mathbf{K}}\|\\
&\quad\le L^2|z|^2\|\mathbf{T}\|^{\ell+1}v^{-6}n^{-1}\mathbf{E}|\mathrm{tr}\mathbf{T}\mathcal{B}_{k,n}^{-1}-\mathbf{E}\mathrm{tr}\mathbf{T}\mathcal{B}_n^{-1}|\\
&\quad\le L^2|z|^2\|\mathbf{T}\|^{\ell+1}v^{-6}n^{-1}\Big(\mathbf{E}|\mathrm{tr}\mathbf{T}\mathcal{B}_{k,n}^{-1}-\mathrm{tr}\mathbf{T}\mathcal{B}_n^{-1}| &(3.9)\\
&\qquad+\mathbf{E}|\mathrm{tr}\mathbf{T}\mathcal{B}_n^{-1}-\mathbf{E}\mathrm{tr}\mathbf{T}\mathcal{B}_n^{-1}|\Big) &(3.10)\\
&\quad\to 0 &(3.11)
\end{aligned}$$

where, to obtain the limit, the expectation in (3.9) can be estimated by Lemma 2.6 of Silverstein and Bai (1995), and the expectation in (3.10) can be estimated by the similar martingale decomposition in Step 1. Hence, we finally reached

$$\frac{1}{p}\Big(\mathbf{E}\mathrm{tr}\mathbf{T}^{\ell}(\widetilde{\mathbf{K}}-z\mathbf{I})^{-1}-\mathbf{E}\mathrm{tr}\mathbf{T}^{\ell}\mathcal{B}_n^{-1}\Big)\to 0,$$

as $n \to \infty$. Noting (3.8), we have

$$\frac{1}{p}\mathbf{E}\mathrm{tr}\Big(\frac{\mathbf{T}}{1+c_n a_n(z)} - z\mathbf{I}\Big)^{-1} - \mathbf{E}m_n(z) \to 0, \tag{3.12}$$

$$\frac{1}{p}\mathbf{E}\mathrm{tr}\mathbf{T}\Big(\frac{\mathbf{T}}{1+c_n a_n(z)} - z\mathbf{I}\Big)^{-1} - a_n(z) \to 0, \tag{3.13}$$

where $a_n(z) = p^{-1}\mathbf{E}\mathrm{tr}\mathbf{T}\mathcal{B}_n^{-1}$. After noticing $|(1+c_n a_n(z))^{-1}| \le |z|/v$, multiply $(1+c_n a_n(z))^{-1}$ on both sides of (3.13) to obtain

$$1 + \frac{z}{p}\mathbf{E}\mathrm{tr}\Big(\frac{\mathbf{T}}{1+c_n a_n(z)} - z\mathbf{I}\Big)^{-1} - \frac{a_n(z)}{1+c_n a_n(z)} \to 0,$$

as $n \to \infty$. Then, by (3.12), we have

$$1 + z\mathbf{E}m_n(z) - \frac{a_n(z)}{1+c_n a_n(z)} \to 0.$$

From this, we conclude

$$\frac{1}{1+c_n a_n(z)} = 1 - c_n(1 + z\mathbf{E}m_n(z)) + o(1).$$

Substituting this into (3.12), we obtain

$$\frac{1}{p}\mathbf{E}\mathrm{tr}\Big(\mathbf{T}(1 - c_n(1+z\mathbf{E}m_n(z))) - z\mathbf{I}\Big)^{-1} - \mathbf{E}m_n(z) \to 0. \tag{3.14}$$

For each fixed z, $\{\mathbf{E}m_n(z)\}$ is a bounded sequence. Thus, for any subsequence n', there is a subsequence $\{n''\}$ of $\{n'\}$ such that $\mathbf{E}m_{n''}(z)$ tends to a limit, say m. Then m should satisfy the equation

$$m = \int \frac{1}{t(1-c-czm) - z} dH(t). \tag{3.15}$$

Since $\Im(\mathbf{E}m_n(z)) > 0$, we conclude that $\Im(m) \ge 0$. Obviously, it is impossible that $\Im(m) = 0$ because the right hand side of (3.15) has a positive imaginary part.

By the uniqueness of the solution to the equation (3.15), proved in the next step, we conclude that for all z with $\Im(z) > 0$, $\mathbf{E}m_n(z)$ converges to a limit which is the unique solution to (3.15).

By Step 1, we conclude that for each z with $\Im(z) > 0$, $m_n(z) \to m(z)$ a.s. as $n \to \infty$. Finally, applying the Vitali Lemma, we conclude that, with probability 1 on any compact subset of $\{z : \Im(z) > 0\}$, $m_n(z) \to m(z)$, which satisfies the equation (3.15).

Step 3. Uniqueness of solution of (1.1).

Suppose we have two solutions $m_1,\ m_2 \in \mathcal{C}^+$ of the equation (1.1). Let $M_1(z) = -(1-c)/z + cm_1(z)$ and $M_2(z) = -(1-c)/z + cm_2(z)$. Both $M_1(z)$ and $M_2(z)$ are Stieltjes transforms of some distributions. Hence $\Im(M_1(z)) > 0$ and $\Im(M_2(z)) > 0$. Since $m_1(z)$ and $m_2(z)$ are solutions of (1.1), we have

$$M_1 = \Big(c\int \frac{t}{tM_1+c}dH(t) - z\Big)^{-1}, \tag{3.16}$$

$$M_2 = \Big(c\int \frac{t}{tM_2+c}dH(t) - z\Big)^{-1}. \tag{3.17}$$

Hence

$$\begin{aligned} M_1 - M_2 = c(M_1-M_2)&\int \frac{t^2}{(tM_1+c)(tM_2+c)}dH(t) \\ &\Big(c\int \frac{t}{tM_1+c}dH(t) - z\Big)^{-1}\Big(c\int \frac{t}{tM_2+c}dH(t) - z\Big)^{-1}. \end{aligned}$$

If $m_1 \neq m_2$, we have

$$\begin{aligned} 1 = c&\int \frac{t^2}{(tM_1+c)(tM_2+c)}dH(t) \\ &\times\Big(c\int \frac{t}{tM_1+c}dH(t) - z\Big)^{-1}\Big(c\int \frac{t}{tM_2+c}dH(t) - z\Big)^{-1} \end{aligned}$$

which, by the Cauchy inequality yields

$$\begin{aligned} 1 \le c&\Big(\int \frac{t^2}{|tM_1+c|^2}dH(t)\int \frac{t^2}{|tM_2+c|^2}dH(t)\Big)^{\frac{1}{2}} \\ &\Big|\Big(c\int \frac{t}{tM_1+c}dH(t) - z\Big)^{-1}\Big(c\int \frac{t}{tM_2+c}dH(t) - z\Big)^{-1}\Big|. \end{aligned} \tag{3.18}$$

From (3.16) and (3.17), we have

$$\begin{aligned} \Im\ M_j &= \Big(v + c\ \Im M_j\int \frac{t^2}{|tM_j+c|^2}dH(t)\Big)\Big|\Big(c\int \frac{t}{tM_j+c}dH(t) - z\Big)\Big|^{-2} \\ &> c\Im M_j\int \frac{t^2}{|tM_j+c|^2}dH(t)\Big|\Big(c\int \frac{t}{tM_j+c}dH(t) - z\Big)\Big|^{-2}, \end{aligned}$$

which implies that for both $j = 1$ and 2,

$$1 > c\int \frac{t^2}{|tM_j+c|^2}dH(t)\Big|\Big(c\int \frac{t}{tM_j+c}dH(t) - z\Big)\Big|^{-2}. \tag{3.19}$$

The contradiction of (3.18) and (3.19) proves that $m_1 = m_2$ and hence (1.1) has at most one solution. The existence of solutions to (1.1) has been seen in Step 2, the proof is complete.

Proof of Corollary 1.1. Under the conditions of Corollary 1.1, for any nonrandom $p \times p$ matrix $B = (b_{jk})$ with bounded norm, we have

$$\begin{aligned}
&n^{-2}\mathbf{E}\Big|\sum_{j,l} b_{j,l}(\bar{X}_{jk}X_{lk} - t_{jl})\Big|^2 \\
&\quad = n^{-2}\mathbf{E}\sum_{j,l}\sum_{j',l'} b_{j,l}\bar{b}_{j',l'}\mathbf{E}(\bar{X}_{jk}X_{lk} - t_{jl})(X_{j'k}\bar{X}_{l'k} - \bar{t}_{j'l'}) \\
&\quad \le n^{-2}\mathbf{E}(\sum_{\Lambda}|b_{j,l}\bar{b}_{j',l'}|^2)^{\frac{1}{2}}\Big(\sum_{\Lambda}\Big|\mathbf{E}(\bar{X}_{jk}X_{lk} - t_{jl})(X_{j'k}\bar{X}_{l'k} - \bar{t}_{j'l'})\Big|^2\Big)^{\frac{1}{2}} \\
&\qquad +2n^{-1}\max_{j\neq l}\mathbf{E}|\bar{X}_{jk}X_{lk} - t_{jl}|^2\mathbf{E}\|(b_{j,l})\|^2 \\
&\quad \le n^{-1}\mathbf{E}\|(b_{j,l})\|^2\Big(\sum_{\Lambda}\Big|\mathbf{E}(\bar{X}_{jk}X_{lk} - t_{jl})(X_{j'k}\bar{X}_{l'k} - \bar{t}_{j'l'})\Big|^2\Big)^{\frac{1}{2}} \\
&\qquad +2n^{-1}\max_{j\neq l}\mathbf{E}\Big|\bar{X}_{jk}X_{lk} - t_{jl}\Big|^2\mathbf{E}\|(b_{j,l})\|^2 \\
&\quad \to 0
\end{aligned}$$

where in the above inequality, we used

$$\sum_{j,l}\sum_{j',l'}|b_{j,l}\bar{b}_{j',l'}| = (\sum_{j,l}|b_{j,l}|)^2 \le n^2\|(b_{j,l})\|^4.$$

Acknowledgement

The authors would like to thank two referees for their valuable comments. The research of Zhidong Bai is partially supported by NSFC 10571020 and NUS grant R-155-000-056-112, and that of W. Zhou is partially supported by R-155-000-076-112 at the National University of Singapore.

References

Bai, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner Matrices. *Ann. Probab.* **21**, 625-648.

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

Bai, Z. D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. *Ann. Probab.* **26**, 316-345.

Bai, Z. D. and Silverstein, J. W. (1999). Exact separation of eigenvalues of large dimensional sample covariance matrices. *Ann. Probab.* **27**, 1536-1555.

Bai, Z. D. and Silverstein, J. W. (2004). Clt for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32**, 553-605.

Bai, Z. D., Yin, Y. Q. and Krishnaiah, P. R. (1986). On limiting spectral distribution of product of two random matrices when the underlying distribution is isotropic. *J. Multivariate Anal.* **19**, 189-200.

Giné, E., Götze, F. and Mason, D. M. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* **25**, 1514-1531.

Hájek, J., Šidák, Z. and Sen, P. K. (1999). *Theory of Rank Tests.* Academic Press, New York.

Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods.* Wiley, New York.

Jiang, T. (2004). The limiting distributions of eigenvalues of sample correlation matrices. Sankhya **66**, 35-48.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.* **72**, 507-536.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **55**, 331-339.

Silverstein, J. W. 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.

Spearman, C. (1904). The proof and measurement of association between two things. *Amer. J. Psychol.* **15**, 72-101.

Wachter, Kenneth W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6**, 1-18.

Yin, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20**, 50-68.

Yin, Y. Q. and Krishnaiah, P. R. (1986). Limit theorems for the eigenvalues of product of large-dimensional random matrices when the underlying distribution is isotropic. *Teor. Veroyatnost. i Primenen.* **31**, 394-398.

School of Mathematics & Statistics, Northeast Normal University, Changchun, 130024, PRC.

E-mail: baizd@nenu.edu.cn

Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546.

E-mail: stazw@nus.edu.sg

(Received July 2006; accepted September 2006)