Submitted to the Annals of Applied Probability
arXiv: math.PR/0567508

LIMITING SPECTRAL DISTRIBUTION OF A SYMMETRIZED AUTO-CROSS COVARIANCE MATRIX

By Baisuo Jin*,§, Chen Wang¶ Z. D. Bai†,¶,‖
K. Krishnan Nair** and Matthew Harding‡,**

University of Science and Technology of China§, National University of Singapore,¶ Northeast Normal University, ‖ *and Stanford University***

This paper studies the limiting spectral distribution (LSD) of a symmetrized auto-cross covariance matrix. The auto-cross covariance matrix is defined as $\mathbf{M}_\tau = \frac{1}{2T}\sum_{j=1}^{T}(\mathbf{e}_j\mathbf{e}_{j+\tau}^* + \mathbf{e}_{j+\tau}\mathbf{e}_j^*)$, where $\mathbf{e}_j$ is an N dimensional vectors of independent standard complex components with properties stated in Theorem (1.1) and τ is the lag. $\mathbf{M}_0$ is well studied in the literature whose LSD is the Marčenko-Pastur (MP) Law. The contribution of this paper is in determining the LSD of $\mathbf{M}_\tau$ where $\tau \geq 1$. It should be noted that the LSD of the $\mathbf{M}_\tau$ does not depend on τ. This study was arising from the investigation and plays an key role in the model selection of any large dimensional model with a lagged time series structure which are central to large dimensional factor models and singular spectrum analysis.

*Research of this author was supported by NSF China Young Scientist Grant 11101397

†Research of this author was supported by NSF China 11171057 as well as by Program for Changjiang Scholars and Innovative Research Team in University

‡The research of this author was supported by Stanford Presidential Fund for Innovation in International Studies

AMS 2000 subject classifications: Primary 60F15, 15A52, 62H25; secondary 60F05, 60F17

Keywords and phrases: Auto-cross covariance, Factor analysis, Mačhenko-Pastur law, Limiting spectral distribution, Order detection, Random matrix theory, Stieltjes transform

1. Introduction. Over the last decade and as a result of new sources of large data, the analysis of high dimensional statistical models has received renewed attention. These models are currently being analyzed within the context of Random Matrix Theory (RMT) in many areas such as statistics (Bai & Silverstein, 2010), economics (Harding, 2012; Onatski, 2009, 2012) and engineering (Rao & Edelman, 2008; Tulino & Verdu, 2004). The asymptotic framework assumes that both the dimension corresponding to the number of individual units, N, and the number of samples T are large.

Suppose $\mathbf{A}_n$ is an $n\times n$ random Hermitian matrix with eigenvalues $\lambda_j, j = 1, 2, \cdots, n$. Define a one-dimensional distribution function of the eigenvalues

$$F^{\mathbf{A}_n}(x) = \frac{1}{n}\sharp\{j \leq n : \lambda_j \leq x\}$$

and $F^{\mathbf{A}_n}(x)$ is called the empirical spectral distribution (ESD) of matrix $\mathbf{A}_n$. Here $\sharp E$ denotes the cardinality of the set E. The limit distribution of $\{F^{\mathbf{A}_n}\}$ for a given sequence of random matrices $\{\mathbf{A}_n\}$ is called the limiting spectral distribution (LSD). For any function of bounded variation G, the *Stieltjes transform* of G is defined as $m_G(z) \equiv \int \frac{1}{\lambda - z} dG(\lambda)$ where $z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im z > 0\}$. For any $n \times n$ matrix $\mathbf{A}_n$ with real eigenvalues λ_1,...,λ_n, the Stieltjes transform of $F^{\mathbf{A}_n}$ is

$$m_{F^{\mathbf{A}_n}}(z) = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \tfrac{1}{\lambda_i - z} = \tfrac{1}{n}\mathrm{tr}(\mathbf{A}_n - z\mathbf{I})^{-1}.$$

Similar to the Fourier transformation in probability theory, there is also a one to one correspondence between the distributions and their Stieltjes transforms via the *inversion formula*: for any function of bounded variation G which is continuous at a and b,

$$G\{[a, b]\} = \tfrac{1}{\pi} \lim_{\eta \to 0^+} \textstyle\int_a^b \Im m_G(\xi + i\eta) d\xi.$$

Besides, the continuity theorem holds, that is, a sequence of distributions tends to a weak limit if and only if their Stieltjes transforms tends to that of the limiting distribution. Therefore, to find the limiting distribution, one can work on finding the limiting Stieltjes transform and use the inversion formula to obtain the limiting distribution.

Research on the LSD of large dimensional random matrices dates back to the Wigner (1955,1958). In these studies, he established that the ESD of a large dimensional Wigner matrix tends to the so-called semicircular law. The LSD of large dimensional sample covariance matrices was studied by Marčenko and Pastur (1967) and the limiting distribution is referred to as the MP law. Further research efforts were conducted to estimate the LSD of a product of two random matrices. To this end, pioneering work was done by Wachter (1980), who considered the LSD of the multivariate F-matrix, the explicit form of which was derived by Bai, Yin and Krishnaiah(1986) and Silverstein(1995). The existence of the LSD of the matrix sequence $\{\mathbf{S}_n\mathbf{T}_n\}$ was established by Yin and Krishnaiah (1983) where $\mathbf{S}_n$ is a standard Wishart matrix and $\mathbf{T}_n$ is a positive definite matrix. Bai, Miao and Jin (2007) proved the existence of the LSD of $\{\mathbf{S}_n\mathbf{T}_n\}$ where $\mathbf{S}_n$ is a sample covariance matrix and $\mathbf{T}_n$ is an arbitrary Hermitian matrix. In particular, Bai, Miao and Jin (2007) established the explicit form of LSD of $\{\mathbf{S}_n\mathbf{T}_n\}$ where $\mathbf{S}_n$ is a sample covariance matrix and $\mathbf{T}_n$ is Wigner matrix. Random matrices of the form $\mathbf{A}_n + \mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n$ where $\mathbf{A}_n$ is Hermitian matrix, $\mathbf{T}_n$ is diagonal and $\mathbf{X}_n$ consists of iid (independently and identically distributed) entries, was extensively investigated by many researchers, including Marčenko and Pastur (1967), Grenander and Silverstein (1977), Wachter (1978), Jonsson

(1982), and Silverstein and Bai (1995). Furthermore the LSD of a circulant random matrix was derived by Bose and Mitra (2002) and the LSD of sample correlation matrices was studied by Jiang (2004). Bai and Zhou (2008) considered the LSD of a large-dimensional sample matrix where the assumption of column independence has been relaxed. A large-dimensional vector autoregressive moving average models (LDVARMA) is a special case of the random matrix framework considered by Bai and Zhou (2008). Jin et al. (2009) established the explicit forms of the LSD of covariance matrices of LDVAR(1) and LDVMA(1). Wang et al. (2011) established the relationship between the power spectral density function and LSD of covariance matrices of LDVARMA(p,q). A detailed exposition of spectral properties of random matrices is presented in Bai and Silverstein (2004, 2010) and Zheng (2012).

1.1. *Motivation and Main Result.* In this paper, we will focus our attention on the LSD of a symmetrized auto-cross covariance matrix $\mathbf{M}_\tau = \sum_{k=1}^{T}(\boldsymbol{\gamma}_k\boldsymbol{\gamma}_{k+\tau}^* + \boldsymbol{\gamma}_{k+\tau}\boldsymbol{\gamma}_k^*)$. where $\boldsymbol{\gamma}_k = \frac{1}{\sqrt{2T}}\mathbf{e}_k$, $\mathbf{e}_k = (\varepsilon_{1k}, \cdots, \varepsilon_{Nk})'$ and $\{\varepsilon_{it}\}$ are independent random variables with mean 0 and variance σ^2. Here, $\tau \geq 1$ denotes the number of lags. The motivation of this paper comes from any large dimensional model with a lagged time series structure which are central to large dimensional dynamic factor models (Forni & Lippi, 2001) and singular spectrum analysis (Vautard et al., 1992; Zhigljavsky, 2012).

Consider the framework of a large dimensional dynamic k-factor model with lag q to understand the underlying motivation of this work. This takes the following form

$$\mathbf{R}_t = \sum_{i=0}^{q} \boldsymbol{\Lambda}_i \mathbf{F}_{t-i} + \mathbf{e}_t, t = 1, ..., T$$

where $\mathbf{\Lambda}_i$'s are $N \times k$ non-random matrices with full rank. For $t = 1, ..., T$, F_t's are k-dimensional vectors of iid standard complex components and $\mathbf{e}_t$'s are N-dimensional vectors of iid complex components with mean zero and finite second moment σ^2, independent of $\mathbf{F}_t$. This model can be viewed as a large dimensional *information-plus-noise* type model (Dozier and Silverstein (2007a, 2007b); Bai and Silverstein (2012)), with information contained in the summation part and noise in $\mathbf{e}_t$'s. Here "large dimension" refers to N and T, while the number of factors k and the number of lags q are small and fixed. In some applications, especially those based on high-frequency observations, it may also be interesting to consider the case where the number of lags is large or the data exhibits long-range dependency. Under this high dimensional setting, an important statistical problem is the estimation of k and q (Bai & Ng, 2002; Harding, 2012). Let τ be a nonnegative integer. For $j = 1, ..., T$, define

$$\mathbf{\Phi}(\tau) = \frac{1}{2T} \sum_{j=1}^{T} (\mathbf{R}_j \mathbf{R}_{j+\tau}^* + \mathbf{R}_{j+\tau} \mathbf{R}_j^*)$$

and

$$\mathbf{M}_\tau = \textstyle\sum_{j=1}^{T} (\boldsymbol{\gamma}_j \boldsymbol{\gamma}_{j+\tau}^* + \boldsymbol{\gamma}_{j+\tau} \boldsymbol{\gamma}_j^*), \text{ where } \boldsymbol{\gamma}_j = \frac{1}{\sqrt{2T}} \mathbf{e}_j.$$

Note that essentially, $\mathbf{M}_\tau$ and $\mathbf{\Phi}(\tau)$ are symmetrized auto-cross covariance matrices at lag τ and generalize the usual sample covariance matrices $\mathbf{M}_0$ and $\mathbf{\Phi}(0)$. The matrix $\mathbf{M}_0$ is well studied in the literature and it is well known that the limiting spectral distribution (LSD) has an MP law (Marčenko and Pastur, 1967). Moreover, when $\tau = 0$ and assuming that $\text{Cov}(\mathbf{F}_t) = \mathbf{\Sigma}_f = \mathbf{I}_{k(q+1)}$, the population covariance matrix of $\mathbf{R}_t$ has the same eigenvalues as those of

$$\begin{pmatrix} \sigma^2\mathbf{I} + \mathbf{\Lambda}^*\mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I} \end{pmatrix}$$

where $\mathbf{\Lambda} = (\mathbf{\Lambda}_0, \mathbf{\Lambda}_1, \cdots, \mathbf{\Lambda}_q)$ whose dimension is $N \times k(q+1)$ and the matrix at the right-lower corner is $(N-k(q+1)) \times (N-k(q+1))$, respectively. Therefore, we have the *spiked population model framework* (Johnstone, 2001), Baik and Silverstein (2006), Bai and Yao (2008) . In fact, under certain conditions, we can estimate $k(q+1)$ by counting the number of eigenvalues of $\mathbf{\Phi}(0)$ that are larger than $\sigma^2(1+\sqrt{c})^2$, where c is the limiting ratio of N/T. However, to estimate the values of k and q separately, we need to study the LSD of $\mathbf{M}_\tau$ for at least one $\tau \geq 1$.

It is interesting to note that for $\tau \geq 1$ (τ being a fixed integer), the LSD of $\mathbf{M}_\tau$ does not depend on τ (see Theorem 1.1 for details). However the number of eigenvalues of $\mathbf{\Phi}(\tau)$ that lie outside the support of the LSD of $\mathbf{M}_\tau$ at lags $1 \leq \tau \leq q$ is dependent on the lag τ; and is different with those obtained at lags $\tau > q$. This is mainly because of the contribution of eigenvalues of the terms containing factor and error components are non-zero for $\tau \geq 1$. Thus, we can separate the estimates of k and q by counting the number of eigenvalues of $\mathbf{\Phi}(\tau)$ that lie outside the support of the LSD of $\mathbf{M}_\tau$ from $\tau = 0, 1, 2, \cdots, q, q+1, \cdots$.

Unlike the case $\tau = 0$, not much is known in the literature for $\mathbf{M}_\tau$ as $\tau \geq 1$. The goal of this paper is to derive the LSD of $\mathbf{M}_\tau$ denoted as F_τ i.e. $F_\tau(x) = \lim_{N\to\infty} F^{\mathbf{M}_\tau}(x)$. Here $F^{\mathbf{A}}$ denotes the ESD of $\mathbf{A}$. In our derivation of the LSD of $\mathbf{M}_\tau$, a recursive method is created to solve a disturbed difference equations of order 2.

The main result of this paper is the following theorem.

THEOREM 1.1. *Assume:*

(a) $\tau \geq 1$ *is a fixed integer.*

(b) $\mathbf{e}_k = (\varepsilon_{1k}, \cdots, \varepsilon_{Nk})'$, $k = 1, 2, ..., T+\tau$, *are* N *dimensional vectors of independent standard complex components with* $\sup_{1\leq i\leq N, 1\leq t\leq T+\tau} E|\varepsilon_{it}|^{2+\delta} \leq M < \infty$ *for some* $\delta \in (0, 2]$, *and for any* $\eta > 0$,

$$\frac{1}{\eta^{2+\delta}NT}\sum_{i=1}^{N}\sum_{t=1}^{T+\tau} E(|\varepsilon_{it}|^{2+\delta} I(|\varepsilon_{it}| \geq \eta T^{1/(2+\delta)})) = o(1). \tag{1.1}$$

(c) $N/(T+\tau) \to c > 0$ *as* $N, T \to \infty$.

(d) $\mathbf{M}_\tau = \sum_{k=1}^{T}(\boldsymbol{\gamma}_k\boldsymbol{\gamma}_{k+\tau}^* + \boldsymbol{\gamma}_{k+\tau}\boldsymbol{\gamma}_k^*)$, *where* $\boldsymbol{\gamma}_k = \frac{1}{\sqrt{2T}}\mathbf{e}_k$.

Then as $N, T \to \infty$, $F^{\mathbf{M}_\tau} \xrightarrow{D} F_\tau$ *a.s. and* F_τ *has a density function given by*

$$\phi_c(x) = \frac{1}{2c\pi}\sqrt{\frac{y_0^2}{1+y_0} - \left(\frac{1-c}{|x|} + \frac{1}{\sqrt{1+y_0}}\right)^2},\ |x| \leq a,$$

where

$$a = \begin{cases} \frac{(1-c)\sqrt{1+y_1}}{y_1-1}, & c \neq 1, \\ 2, & c = 1, \end{cases}$$

y_0 *is the largest real root of the equation:* $y^3 - \frac{(1-c)^2-x^2}{x^2}y^2 - \frac{4}{x^2}y - \frac{4}{x^2} = 0$ *and* y_1 *is the only real root of the equation:*

$$((1-c)^2 - 1)y^3 + y^2 + y - 1 = 0 \tag{1.2}$$

such that $y_1 > 1$ *if* $c < 1$ *and* $y_1 \in (0, 1)$ *if* $c > 1$. *Further, if* $c > 1$, *then* F_τ *has a point mass* $1 - 1/c$ *at the origin.*

Remark 1.1. *Notice that as long as* $\tau \geq 1$, F_τ *is the same as* τ *takes different values. However,* F_τ *is different from and can not be reduced to* F_0, *the distribution of MP law.*

Remark 1.2. *When $\tau = o(T)$, the conclusion still holds (see remark after the proof of Lemma B.3 for details). When $\frac{\tau}{T} \to d$ for some $d > 0$, we conjecture that the LSD will depend on d as well and leave it as future research.*

Figure 1 displays the density functions $\phi_c(x)$ with $c = 0.2, 0.5$ and 0.7. Figure 2 displays the density functions $\phi_c(x)$ with $c = 1.5, 2$ and 2.5. It is shown from these two figures that as c increases, the support of $\phi_c(x)$ becomes wider, and $\phi_c(x)$ has the maximum at $x = 0$ which is sharper as c gets closer to 1.

FIG 1. *Density functions $\phi_c(x)$ of the LSD of $\mathbf{M}_\tau$ with $c = 0.2$ (the solid line), $c = 0.5$ (the dashed line) and $c = 0.7$ (the dotted line).*

The rest of this paper is organized as follows. The truncation and centralization steps are provided in Section 2. Section 3 outlines the proof of the main theorem Theorem 1.1. Justification of variable truncation, central-

FIG 2. *Density functions $\phi_c(x)$ of the LSD of $\mathbf{M}_\tau$ with $c = 1.5$ (the solid line), $c = 2$ (the dashed line) and $c = 2.5$ (the dotted line). Note that the area under each density function curve is $1/c$.*

ization and standardization is provided in Appendix A and some technical lemmas used for the derivation of Theorem 1.1 are presented in Appendix B.

2. Truncation, Centralization and Standardization. First, we can select a sequence $\eta_N \downarrow 0$ such that (1.1) remains true when η is replaced by η_N.

After truncation at $\eta_N T^{1/(2+\delta)}$, centralization and standardization, in what follows, we may assume that

$$|\varepsilon_{ij}| \le \eta_N T^{1/(2+\delta)},\ \mathrm{E}\varepsilon_{ij} = 0,\ \mathrm{E}|\varepsilon_{ij}|^2 = 1,\ \mathrm{E}|\varepsilon_{ij}|^{2+\delta} < M.$$

The details of verification is provided in Appendix A.

3. Derivation of the LSD of $\mathbf{M}_\tau$. In this section, we will provide the proof for the derivation of Theorem 1.1. To this end, we start with a section on notation followed by the proof.

3.1. *Notation .* Let the Stieltjes transform of $\mathbf{M}_\tau$ be denoted by $m_N(z) = \frac{1}{N}\mathrm{tr}(\mathbf{M}_\tau - z\mathbf{I}_N)^{-1}$ where $z = u + iv$, $v > 0$. We shall prove that $m_N(z) \to m(z)$ for some $m(z)$. It follows that the LSD of $\mathbf{M}_\tau$ exists and has a probability density function $\lim_{v\to 0}\frac{1}{\pi}\Im(m(x+iv))$.
Define

$$\begin{aligned}
\mathbf{A} &= \mathbf{M}_\tau - z\mathbf{I}_N,\\
\mathbf{A}_k &= \mathbf{A} - \boldsymbol{\gamma}_k(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau})^* - (\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau})\boldsymbol{\gamma}_k^*,\\
\mathbf{A}_{k,k+\tau,\cdots,k+n\tau} &= \mathbf{A}_{k,k+\tau,\cdots,k+(n-1)\tau} - \boldsymbol{\gamma}_{k+(n+1)\tau}\boldsymbol{\gamma}_{k+n\tau}^* - \boldsymbol{\gamma}_{k+n\tau}\boldsymbol{\gamma}_{k+(n+1)\tau}^*, n \geq 1,
\end{aligned}$$

for $k \in [\tau+1, T]$. Note that $\mathbf{A}_{k,k+\tau,\cdots,k+n\tau}$ is independent of $\boldsymbol{\gamma}_k, \cdots, \boldsymbol{\gamma}_{k+n\tau}$. For $k \leq \tau$ or $k > T$, we still use the definition of $\mathbf{A}_k$ with the convention that $\boldsymbol{\gamma}_l = \mathbf{0}$ for $l \leq 0$ or $l > T + \tau$.

3.2. *Derivation.* By

$$\mathbf{A} = \sum_{k=1}^{T}(\boldsymbol{\gamma}_k\boldsymbol{\gamma}_{k+\tau}^* + \boldsymbol{\gamma}_{k+\tau}\boldsymbol{\gamma}_k^*) - z\mathbf{I}_N$$

we have

$$\mathbf{I}_N = \sum_{k=1}^{T}(\mathbf{A}^{-1}\boldsymbol{\gamma}_k\boldsymbol{\gamma}_{k+\tau}^* + \mathbf{A}^{-1}\boldsymbol{\gamma}_{k+\tau}\boldsymbol{\gamma}_k^*) - z\mathbf{A}^{-1}.$$

Taking trace and dividing by N, we obtain

$$1 + zm_n(z) = \frac{1}{N}\sum_{k=1}^{T}(\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}^{-1}\boldsymbol{\gamma}_k + \boldsymbol{\gamma}_k^*\mathbf{A}^{-1}\boldsymbol{\gamma}_{k+\tau}). \tag{3.1}$$

Taking expectation on both sides, we obtain

$$1 + z\mathrm{E}m_n(z) = \frac{1}{N}\sum_{k=1}^{T}(\mathrm{E}\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}^{-1}\boldsymbol{\gamma}_k + \mathrm{E}\boldsymbol{\gamma}_k^*\mathbf{A}^{-1}\boldsymbol{\gamma}_{k+\tau}). \tag{3.2}$$

Applying the identity

$$(\mathbf{B} + \alpha\boldsymbol{\gamma}^*)^{-1} = \mathbf{B}^{-1} - \frac{\mathbf{B}^{-1}\alpha\boldsymbol{\gamma}^*\mathbf{B}^{-1}}{1 + \boldsymbol{\gamma}^*\mathbf{B}^{-1}\alpha}, \tag{3.3}$$

for any nonsingular matrix $\mathbf{B}$, we have

$$\begin{aligned}\boldsymbol{\gamma}_k^*\mathbf{A}^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau}) &= \frac{\boldsymbol{\gamma}_k^*\tilde{\mathbf{A}}_k^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau})}{1 + \boldsymbol{\gamma}_k^*\tilde{\mathbf{A}}_k^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau})} \\ &= 1 - \frac{1}{1 + \boldsymbol{\gamma}_k^*\tilde{\mathbf{A}}_k^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau})},\end{aligned}$$

where $\tilde{\mathbf{A}}_k = \mathbf{A} - (\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau})\boldsymbol{\gamma}_k^*$ and we have used the previously made convention that $\boldsymbol{\gamma}_l = \mathbf{0}$ for $l \leq 0$ or $l > T + \tau$. Note that $\mathbf{A}_k = \tilde{\mathbf{A}}_k - \boldsymbol{\gamma}_k(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau})^*$. Using (3.3) again, we have

$$\boldsymbol{\gamma}_k^*\tilde{\mathbf{A}}_k^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau}) = \boldsymbol{\gamma}_k^*\mathbf{A}_k^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau}) - \frac{\boldsymbol{\gamma}_k^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_k(\boldsymbol{\gamma}_{k+\tau}^* + \boldsymbol{\gamma}_{k-\tau}^*)\mathbf{A}_k^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau})}{1 + (\boldsymbol{\gamma}_{k+\tau}^* + \boldsymbol{\gamma}_{k-\tau}^*)\mathbf{A}_k^{-1}\boldsymbol{\gamma}_k}.$$

By Lemmas B.1 and B.2, we have

$$\boldsymbol{\gamma}_k^*\tilde{\mathbf{A}}_k^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau}) = -\frac{c}{2}m_n(z)(\boldsymbol{\gamma}_{k+\tau}^* + \boldsymbol{\gamma}_{k-\tau}^*)\mathbf{A}_k^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau}) + o_{a.s.}(1).$$

Consequently,

$$\boldsymbol{\gamma}_k^*\mathbf{A}^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau}) = 1 - \frac{1}{1 - \frac{c}{2}m_n(z)(\boldsymbol{\gamma}_{k+\tau}^* + \boldsymbol{\gamma}_{k-\tau}^*)\mathbf{A}_k^{-1}(\boldsymbol{\gamma}_{k+\tau} + \boldsymbol{\gamma}_{k-\tau})} + o_{a.s.}(1).$$

Write $\mathbf{A}_{k,k+\tau} = \mathbf{A}_k - \boldsymbol{\gamma}_{k+2\tau}\boldsymbol{\gamma}_{k+\tau}^* - \boldsymbol{\gamma}_{k+\tau}\boldsymbol{\gamma}_{k+2\tau}^*$ which is independent of

$\boldsymbol{\gamma}_{k+\tau}$. Then, using (3.3) again, we obtain

$$\begin{aligned}
\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_{k+\tau} &= \frac{\boldsymbol{\gamma}_{k+\tau}^*(\mathbf{A}_{k,k+\tau}+\boldsymbol{\gamma}_{k+2\tau}\boldsymbol{\gamma}_{k+\tau}^*)^{-1}\boldsymbol{\gamma}_{k+\tau}}{1+\boldsymbol{\gamma}_{k+2\tau}^*(\mathbf{A}_{k,k+\tau}+\boldsymbol{\gamma}_{k+2\tau}\boldsymbol{\gamma}_{k+\tau}^*)^{-1}\boldsymbol{\gamma}_{k+\tau}} \\
&= \frac{\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+\tau} - \frac{\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+\tau}}{1+\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}}}{1+\boldsymbol{\gamma}_{k+2\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+\tau} - \frac{\boldsymbol{\gamma}_{k+2\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+\tau}}{1+\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}}} \\
&= \frac{\frac{c}{2}m_n(z)}{1-\frac{c}{2}m_n(z)\boldsymbol{\gamma}_{k+2\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}} + o_{a.s.}(1). \qquad (3.4)
\end{aligned}$$

By the same reason,

$$\boldsymbol{\gamma}_{k-\tau}^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_{k-\tau} = \frac{\frac{c}{2}m_n(z)}{1-\frac{c}{2}m_n(z)\boldsymbol{\gamma}_{k-2\tau}^*\mathbf{A}_{k,k-\tau}^{-1}\boldsymbol{\gamma}_{k-2\tau}} + o_{a.s.}(1). \tag{3.5}$$

Next, we consider the cross terms. We have

$$\begin{aligned}
\boldsymbol{\gamma}_{k-\tau}^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_{k+\tau} &= \frac{\boldsymbol{\gamma}_{k-\tau}^*(\mathbf{A}_{k,k+\tau}+\boldsymbol{\gamma}_{k+2\tau}\boldsymbol{\gamma}_{k+\tau}^*)^{-1}\boldsymbol{\gamma}_{k+\tau}}{1+\boldsymbol{\gamma}_{k+2\tau}^*(\mathbf{A}_{k,k+\tau}+\boldsymbol{\gamma}_{k+2\tau}\boldsymbol{\gamma}_{k+\tau}^*)^{-1}\boldsymbol{\gamma}_{k+\tau}} \\
&= \frac{\boldsymbol{\gamma}_{k-\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+\tau} - \frac{\boldsymbol{\gamma}_{k-\tau}\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+\tau}}{1+\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}}}{1+\boldsymbol{\gamma}_{k+2\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+\tau} - \frac{\boldsymbol{\gamma}_{k+2\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+\tau}}{1+\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}}} \\
&= \frac{-\frac{c}{2}m_n(z)\boldsymbol{\gamma}_{k-\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}}{1-\frac{c}{2}m_n(z)\boldsymbol{\gamma}_{k+2\tau}^*\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}} + o_{a.s.}(1). \qquad (3.6)
\end{aligned}$$

Suppose that $m_n(z)$ converges to $m(z)$ along some subsequence $N = n'$, by Lemmas B.3 and B.4, (3.2) will converge to

$$c + czm(z) = 1 - \frac{1}{1-\frac{c^2m^2(z)}{2x_1}}, \tag{3.7}$$

where x_1 is the root of the equation $x^2 = x - \frac{c^2m^2(z)}{4}$ with the larger absolute value. Substituting the expression of x_1, we obtain

$$(1 - c^2m^2(z))(c + czm(z) - 1)^2 = 1. \tag{3.8}$$

This can be further simplified to

$$(cm(z))^4 - \frac{2(1-c)}{z}(cm(z))^3 + \frac{(1-c)^2 - z^2}{z^2}(cm(z))^2 + \frac{2(1-c)}{z}cm(z) + \frac{1-(1-c)^2}{z^2} = 0. \tag{3.9}$$

Now, we shall employ the method developed in Bai et al (2007) to solve the 4th degree polynomial equation and identify the unique solution of the limiting spectral distribution. Rewrite the equation (3.9) as

$$((cm(z))^2 - \frac{(1-c)}{z}cm(z) + \frac{y}{2})^2 = (1+y)(cm(z))^2 - \frac{(1-c)}{z}(y+2)cm(z) + \frac{y^2}{4} - \frac{1-(1-c)^2}{z^2}. \tag{3.10}$$

Let y_0 be the largest real root of the equation:

$$y^3 - \frac{(1-c)^2 - z^2}{z^2}y^2 - \frac{4}{z^2}y - \frac{4}{z^2} = 0.$$

Let $f(y) = y^3 - \frac{(1-c)^2 - z^2}{z^2}y^2 - \frac{4}{z^2}y - \frac{4}{z^2}$. For $f(+\infty) > 0, f(0) < 0$, we have $y_0 > 0$. Further if $z \to 0$, then $y_0 \to \infty$ and $z^2 y_0 \to (1-c)^2$. If we replace y by y_0 in equation (3.10), the solutions to (3.9) will be those to equations

$$(cm(z))^2 - \frac{1-c}{z}cm(z) + \frac{1}{2}y_0 = \pm\sqrt{1+y_0}(cm(z) - \frac{(y_0+2)(1-c)}{2z(1+y_0)}),$$

from which we get four roots:

$$\begin{aligned}
m_1(z) &= \frac{(\frac{1-c}{z}+\sqrt{1+y_0})+\sqrt{(\frac{1-c}{z}-\frac{1}{\sqrt{1+y_0}})^2-\frac{y_0^2}{1+y_0}}}{2c}\\
m_2(z) &= \frac{(\frac{1-c}{z}+\sqrt{1+y_0})-\sqrt{(\frac{1-c}{z}-\frac{1}{\sqrt{1+y_0}})^2-\frac{y_0^2}{1+y_0}}}{2c}\\
m_3(z) &= \frac{(\frac{1-c}{z}-\sqrt{1+y_0})+\sqrt{(\frac{1-c}{z}+\frac{1}{\sqrt{1+y_0}})^2-\frac{y_0^2}{1+y_0}}}{2c}\\
m_4(z) &= \frac{(\frac{1-c}{z}-\sqrt{1+y_0})-\sqrt{(\frac{1-c}{z}+\frac{1}{\sqrt{1+y_0}})^2-\frac{y_0^2}{1+y_0}}}{2c}.
\end{aligned}$$

Now, we claim that the density of F_τ at the origin $\lim_{z\to 0} -zm(z)$ satisfies

$$\lim_{z\to 0} -zm(z) = \begin{cases} 1-1/c, & c>1,\\ 0, & c\le 1.\end{cases} \tag{3.11}$$

To show our claim, first by equation (3.8), we have

$$(z^2-c^2z^2m^2(z))(czm(z)-(1-c))^2 = z^2.$$

This means $zm(z)$ must be bounded as $z\to 0$. Otherwise, the LHS of the equation above is unbounded while the RHS tends to 0, which is a contradiction. Hence, the equation above can be simplified as

$$(czm(z))^2(czm(z)-(1-c))^2 = 0.$$

This means there exists a convergent subsequence $\{z_km(z_k)\}$ such that as $\lim_{z_k\to 0} -z_km(z_k)$ can only be either 0 or $1-\frac{1}{c}$. Notice that $\lim_{z\to 0} -zm(z)$ is the density of F_τ at 0 , which is nonnegative. Therefore, as $c<1$, $\lim_{z_k\to 0} -z_km(z_k) \neq 1-\frac{1}{c}$. Hence $\lim_{z_k\to 0} -z_km(z_k) = 0$ and the second part of our claim is proved. When $c \geq 1$, assume $\lim_{z_k \to 0} -z_k m(z_k) \neq 1 - \frac{1}{c}$ i.e. $\lim_{z_k \to 0} -z_k m(z_k) = 0$, then (3.7) becomes

$$c = 1 - \frac{1}{1 - \frac{c^2 m^2(z)}{2x_1}}.$$

Solve this for x_1, and we have

$$x_1 = \frac{c^2 m^2(z)(c-1)}{2c} = \frac{c-2}{2(c-1)}.$$

Here the last equality is due to the fact $1 - c^2 m^2(z) = \frac{1}{(1-c)^2}$ which can be derived from (3.8) and our assumption that $\lim_{z \to 0} -zm(z) = 0$. However, solve the equation $x^2 = x - \frac{c^2 m^2(z)}{4}$ and use the fact $1 - c^2 m^2(z) = \frac{1}{(1-c)^2}$ again, we have

$$x_1 = \frac{1}{2}(1 + \frac{1}{c-1}) = \frac{c}{2(c-1)},$$

which contradicts our last expression of x_1. Hence the first part of the claim is proved.

By (3.11), $\lim_{z \to 0^+} z\sqrt{y} \to |1-c|$, $\lim_{z \to 0^-} z\sqrt{y} \to -|1-c|$ and $\phi_c(x) = \lim_{v \to 0} \frac{1}{\pi} \Im(m(x+iv)) > 0$, we get

$$m(z) = \begin{cases} m_1(z), & z < 0, \\ m_3(z), & z > 0. \end{cases}$$

Therefore, we have

$$\phi_c(x) = \lim_{v \to 0} \frac{1}{\pi} \Im(m(x+iv)) = \frac{1}{2c\pi} \sqrt{\frac{y_0^2}{1+y_0} - (\frac{1-c}{|x|} + \frac{1}{\sqrt{1+y_0}})^2}, |x| \leq a,$$

where y_0 is the largest real root of the equation: $y^3 - \frac{(1-c)^2 - x^2}{x^2} y^2 - \frac{4}{x^2} y - \frac{4}{x^2} = 0$ and a satisfies equations

$$\begin{cases} \frac{y^2}{1+y} - (\frac{1-c}{a} + \frac{1}{\sqrt{1+y}})^2 = 0 \\ y^3 - \frac{(1-c)^2 - a^2}{a^2} y^2 - \frac{4}{a^2} y - \frac{4}{a^2} = 0. \end{cases}$$

Solving these equations under the condition $a > 0$, we have

$$a = \begin{cases} \frac{(1-c)\sqrt{1+y_1}}{y_1-1}, & c \neq 1, \\ 2, & c = 1, \end{cases}$$

where y_1 can be chosen as a real root of the equation:

$$((1-c)^2 - 1)y^3 + y^2 + y - 1 = 0 \tag{3.12}$$

such that $y_1 > 1$ if $c < 1$ and $y_1 \in (0,1)$ if $c > 1$.

To show the unique existence of y_1, let $f(y) = ((1-c)^2 - 1)y^3 + y^2 + y - 1$. If $c < 1$, by $f(-\infty) > 0$, $f(0) < 0$, $f(1) > 0$ and $f(\infty) < 0$, there are three real roots $y_1 > 1$, $y_2 \in (0,1)$ and $y_3 < 0$ of (3.12). Similarly, if $1 < c < 2$, there are three real roots $y_1 \in (0,1)$, $y_2 > 1$ and $y_3 < 0$ of (3.12). If $c = 2$, it is easy to see that there are two real roots of (3.12): $y_1 = (\sqrt{5}-1)/2 \in (0,1)$ and $y_2 = (-\sqrt{5}-1)/2 < 0$. If $c > 2$, by $f(0) < 0$ and $f(1) > 0$, there is a real root $y_1 \in (0,1)$. If there are more than one real roots in the interval $(0,1)$ when $c > 2$, then by the continuity of $f(y)$, the three roots y_1, y_2, y_3 of $f(y)$ are all in the interval $(0,1)$, that would contradict $y_1 + y_2 + y_3 = -1/((1-c)^2 - 1) < 0$. Thus there is only one real root $y_1 \in (0,1)$ if $c > 2$. The proof of Theorem 1.1 is complete.

APPENDIX A: JUSTIFICATION OF TRUNCATION, CENTRALIZATION AND STANDARDIZATION

Note that $\operatorname{rank}(\mathbf{AB} - \mathbf{CD}) \leq \operatorname{rank}(\mathbf{A} - \mathbf{C}) + \operatorname{rank}(\mathbf{B} - \mathbf{D})$, because

$$\mathbf{AB} - \mathbf{CD} = (\mathbf{A} - \mathbf{C})\mathbf{B} + \mathbf{C}(\mathbf{B} - \mathbf{D}).$$

Let $\tilde{\varepsilon}_{it} = \varepsilon_{it} I(|\varepsilon_{it}| < \eta_N T^{1/(2+\delta)})$, $\tilde{\boldsymbol{\gamma}}_k = \frac{1}{\sqrt{2T}}(\tilde{\varepsilon}_{1k}, \cdots, \tilde{\varepsilon}_{Nk})'$ and $\tilde{\mathbf{M}}_\tau = \sum_{k=1}^{T}(\tilde{\boldsymbol{\gamma}}_k \tilde{\boldsymbol{\gamma}}_{k+\tau}^* + \tilde{\boldsymbol{\gamma}}_{k+\tau}\tilde{\boldsymbol{\gamma}}_k^*)$.

By Theorem A.43 of Bai & Silverstein (2010),

$$\begin{aligned}||F^{\mathbf{M}_\tau} - F^{\tilde{\mathbf{M}}_\tau}|| &\le \frac{1}{N}\text{rank}(\mathbf{M}_\tau - \tilde{\mathbf{M}}_\tau)\\ &\le \frac{2}{N}\text{rank}(\sum_{k=1}^{T}(\boldsymbol{\gamma}_k\boldsymbol{\gamma}_{k+\tau}^*) - \sum_{k=1}^{T}(\tilde{\boldsymbol{\gamma}}_k\tilde{\boldsymbol{\gamma}}_{k+\tau}^*))\\ &\le \frac{4}{N}\sum_{i=1}^{N}\sum_{t=1}^{T+\tau} I(|\varepsilon_{it}| \ge \eta_N T^{1/(2+\delta)}).\end{aligned}$$

By (1.1) we have

$$\begin{aligned}&\mathrm{E}(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T+\tau} I(|\varepsilon_{it}| \ge \eta T^{1/(2+\delta)}))\\ \text{(A.1)}\qquad &\le \frac{1}{\eta^{2+\delta}NT}\sum_{i=1}^{N}\sum_{t=1}^{T+\tau}\mathrm{E}(|\varepsilon_{it}|^{(2+\delta)}I(|\varepsilon_{it}| \ge \eta T^{1/(2+\delta)})) = o(1),\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Var}(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T+\tau} I(|\varepsilon_{it}| \ge \eta T^{1/(2+\delta)}))\\ \le\ &\frac{1}{\eta^{2+\delta}N^2T}\sum_{i=1}^{N}\sum_{t=1}^{T+\tau}\mathrm{E}(|\varepsilon_{it}|^{2+\delta}I(|\varepsilon_{it}| \ge \eta T^{1/(2+\delta)})) = o(1/N).\end{aligned}$$

Applying Bernstein's inequality, for all small $\varepsilon > 0$ and large N, we have

$$\mathrm{P}(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T+\tau} I(|\varepsilon_{it}| \ge \eta T^{1/(2+\delta)}) \ge \varepsilon) \le 2e^{-\frac{1}{2}\varepsilon^2 N}.$$

By Borel-Cantelli lemma, with probability 1, we have

$$||F^{\mathbf{M}_\tau} - F^{\tilde{\mathbf{M}}_\tau}|| \to 0.$$

Let $\hat{\varepsilon}_{it} = \tilde{\varepsilon}_{it} - \mathrm{E}\tilde{\varepsilon}_{it}$, $\hat{\boldsymbol{\gamma}}_k = \frac{1}{\sqrt{2T}}(\hat{\varepsilon}_{1k}, \cdots, \hat{\varepsilon}_{Nk})'$, and $\hat{\mathbf{M}}_\tau = \sum_{k=1}^{T}(\hat{\boldsymbol{\gamma}}_k\hat{\boldsymbol{\gamma}}_{k+\tau}^* + \hat{\boldsymbol{\gamma}}_{k+\tau}\hat{\boldsymbol{\gamma}}_k^*)$. By Theorem A.46 of Bai & Silverstein (2010),

$$\begin{aligned}L(F^{\tilde{\mathbf{M}}_\tau} - F^{\hat{\mathbf{M}}_\tau}) &\le \max_k |\lambda_k(\tilde{\mathbf{M}}_\tau) - \lambda_k(\hat{\mathbf{M}}_\tau)| \le ||\tilde{\mathbf{M}}_\tau - \hat{\mathbf{M}}_\tau||\\ &\le 2||\sum_{k=1}^{T}(\hat{\boldsymbol{\gamma}}_k\mathrm{E}\tilde{\boldsymbol{\gamma}}_{k+\tau}^* + \hat{\boldsymbol{\gamma}}_{k+\tau}\mathrm{E}\tilde{\boldsymbol{\gamma}}_k^*)|| + ||\sum_{k=1}^{T}(\mathrm{E}\tilde{\boldsymbol{\gamma}}_k\mathrm{E}\tilde{\boldsymbol{\gamma}}_{k+\tau}^* + \mathrm{E}\tilde{\boldsymbol{\gamma}}_{k+\tau}\mathrm{E}\tilde{\boldsymbol{\gamma}}_k^*)||,\end{aligned}$$

where L is the Levy distance between two distribution functions. For the second part, we have

$$\begin{aligned}
&||\sum_{k=1}^{T}(\mathrm{E}\tilde{\gamma}_k\mathrm{E}\tilde{\gamma}_{k+\tau}^* + \mathrm{E}\tilde{\gamma}_{k+\tau}\mathrm{E}\tilde{\gamma}_k^*)|| \\
\leq\ & \frac{1}{T}\sum_{k=1}^{T}\sum_{i=1}^{N}|\mathrm{E}(\varepsilon_{ik}I(|\varepsilon_{ik}| \geq \eta T^{1/(2+\delta)})\mathrm{E}(\varepsilon_{i(k+\tau)}I(|\varepsilon_{i(k+\tau)}| \geq \eta T^{1/(2+\delta)})| \\
\leq\ & \frac{C}{T^2}\sum_{k=1}^{T+\tau}\sum_{i=1}^{N}\mathrm{E}(|\varepsilon_{ik}|^{2+\delta}I(|\varepsilon_{ik}| \geq \eta T^{1/(2+\delta)}) = o(1).
\end{aligned}$$

For the first part, notice that

$$||\sum_{k=1}^{T}(\hat{\gamma}_k\mathrm{E}\tilde{\gamma}_{k+\tau}^* + \hat{\gamma}_{k+\tau}\mathrm{E}\tilde{\gamma}_k^*)||^2 \leq 2(||\sum_{k=1}^{T}\hat{\gamma}_k\mathrm{E}\tilde{\gamma}_{k+\tau}^*||^2 + ||\sum_{k=1}^{T}\hat{\gamma}_{k+\tau}\mathrm{E}\tilde{\gamma}_k^*||^2).$$

$$\begin{aligned}
||\sum_{k=1}^{T}\hat{\gamma}_k\mathrm{E}\tilde{\gamma}_{k+\tau}^*||^2 \leq\ & \frac{1}{4T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\sum_{k=1}^{T}\hat{\varepsilon}_{ki}\mathrm{E}\tilde{\varepsilon}_{(k+\tau)j})^2 \\
=\ & \frac{1}{4T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k_1=1}^{T}\sum_{k_2=1}^{T}(\hat{\varepsilon}_{k_1i}\hat{\varepsilon}_{k_2i}\mathrm{E}\tilde{\varepsilon}_{(k_1+\tau)j}\mathrm{E}\tilde{\varepsilon}_{(k_2+\tau)j}) \\
=\ & \frac{1}{4T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{k_1=1}^{T}\hat{\varepsilon}_{k_1i}^2(\mathrm{E}\tilde{\varepsilon}_{(k_1+\tau)j})^2 + \sum_{k_1\neq k_2}\hat{\varepsilon}_{k_1i}\hat{\varepsilon}_{k_2i}\mathrm{E}\tilde{\varepsilon}_{(k_1+\tau)j}\mathrm{E}\tilde{\varepsilon}_{(k_2+\tau)j}\right) \\
\equiv\ & J_{11} + J_{12}.
\end{aligned}$$

For $\mathrm{E}\hat{\varepsilon}_{k_1i}^2 < \infty$ and $\mathrm{E}\tilde{\varepsilon}_{(k_1+\tau)j}^{2+\delta} < \infty$, there exist constants C_1, C_2 and C_3 such that

$$\begin{aligned}
\mathrm{E}J_{11} =\ & \frac{1}{4T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k_1=1}^{T}\mathrm{E}\hat{\varepsilon}_{k_1i}^2(\mathrm{E}\tilde{\varepsilon}_{(k_1+\tau)j})^2 \\
\leq\ & \frac{C_1}{4T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k_1=1}^{T}(\mathrm{E}(|\varepsilon_{jk}|I(|\varepsilon_{jk}| \geq \eta T^{1/(2+\delta)})))^2 \\
\leq\ & \frac{C_1}{4T^2\eta^{2(1+\delta)}T^{\frac{2(1+\delta)}{2+\delta}}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k_1=1}^{T}(\mathrm{E}(|\varepsilon_{jk}|^{2+\delta}I(|\varepsilon_{jk}| \geq \eta T^{1/(2+\delta)})))^2 \\
=\ & O(T^{-\frac{\delta}{2+\delta}})
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}J_{11} &= \frac{1}{4^2T^4}\sum_{i=1}^{N}\sum_{k_1=1}^{T}\mathrm{E}(\hat{\varepsilon}_{k_1i}^2-\mathrm{E}\hat{\varepsilon}_{k_1i}^2)^2(\sum_{j=1}^{N}(\mathrm{E}\tilde{\varepsilon}_{(k_1+\tau)j})^2)^2 \\
&\le \frac{C_2}{4^2T^4}\sum_{i=1}^{N}\sum_{k_1=1}^{T}\mathrm{E}\tilde{\varepsilon}_{k_1i}^4(N\frac{1}{\eta^{2(1+\delta)}T^{\frac{2(1+\delta)}{2+\delta}}})^2 \\
&= O(T^{-1-\frac{4\delta}{2+\delta}}).
\end{aligned}$$

The previous two equations imply that $J_{11}\to 0$, a.s..

Furthermore, we have

$$\begin{aligned}
\mathrm{Var}J_{12} &= \frac{1}{4^2T^4}\sum_{i=1}^{N}\sum_{k_1\neq k_2}\mathrm{E}\hat{\varepsilon}_{k_1i}^2\mathrm{E}\hat{\varepsilon}_{k_2i}^2(\sum_{j=1}^{N}\mathrm{E}\tilde{\varepsilon}_{(k_1+\tau)j}\mathrm{E}\tilde{\varepsilon}_{(k_2+\tau)j})^2 \\
&\le \frac{C_3}{4^2T^4}\sum_{i=1}^{N}\sum_{k_1\neq k_2}(\sum_{j=1}^{N}\frac{1}{\eta^{2(1+\delta)}T^{\frac{2(1+\delta)}{2+\delta}}})^2 \\
&= O(T^{-1-\frac{2\delta}{2+\delta}}),
\end{aligned}$$

which implies $J_{12}\to 0$, a.s. Hence, we have $||\sum_{k=1}^{T}\hat{\boldsymbol{\gamma}}_k\mathrm{E}\tilde{\boldsymbol{\gamma}}_{k+\tau}^*||^2\to 0$, a.s.. Similarly $||\sum_{k=1}^{T}\hat{\boldsymbol{\gamma}}_{k+\tau}\mathrm{E}\tilde{\boldsymbol{\gamma}}_k^*||^2\to 0$ a.s. Thus $L(F^{\tilde{\mathbf{M}}_\tau}-F^{\hat{\mathbf{M}}_\tau})\to 0$, a.s..

Now, we want to rescale the variables.

Let $\sigma_{ij}^2=\mathrm{E}|\hat{\varepsilon}_{ij}|^2=\mathrm{E}|\tilde{\varepsilon}_{ij}-\mathrm{E}\tilde{\varepsilon}_{ij}|^2$. Define $E\equiv\{(i,j):\sigma_{ij}^2<1-\Delta\}$ and

$$\check{\varepsilon}_{it}=\begin{cases} X_{it}, & (i,t)\in E, \\ \frac{\hat{\varepsilon}_{it}}{\sigma_{it}}, & otherwise. \end{cases}$$

Here $\Delta=T^{-\frac{\delta}{4+2\delta}}$ and X_{it}'s are iid random variables taking values 1 and -1, each with probability $\frac{1}{2}$. Note that $\mathrm{E}\check{\varepsilon}_{it}=0$ and $\mathrm{Var}(\check{\varepsilon}_{it})=1$.

Let $\check{\boldsymbol{\gamma}}_k=\frac{1}{\sqrt{2T}}(\check{\varepsilon}_{1k},\cdots,\check{\varepsilon}_{Nk})'$ and $\check{\mathbf{M}}_\tau=\sum_{k=1}^{T}(\check{\boldsymbol{\gamma}}_k\check{\boldsymbol{\gamma}}_{k+\tau}^*+\check{\boldsymbol{\gamma}}_{k+\tau}\check{\boldsymbol{\gamma}}_k^*)$. For simplicity, denote $\mathbf{A}=(\hat{\boldsymbol{\gamma}}_1,\hat{\boldsymbol{\gamma}}_2,...,\hat{\boldsymbol{\gamma}}_T)$, $\mathbf{B}=(\hat{\boldsymbol{\gamma}}_{1+\tau},\hat{\boldsymbol{\gamma}}_{2+\tau},...,\hat{\boldsymbol{\gamma}}_{T+\tau})$, $\check{\mathbf{A}}=(\check{\boldsymbol{\gamma}}_1,\check{\boldsymbol{\gamma}}_2,...,\check{\boldsymbol{\gamma}}_T)$ and $\check{\mathbf{B}}=(\check{\boldsymbol{\gamma}}_{1+\tau},\check{\boldsymbol{\gamma}}_{2+\tau},...,\check{\boldsymbol{\gamma}}_{T+\tau})$. Then by Corollary A.41 of

Bai & Silverstein (2010), we have

$$\begin{aligned}
& L^3(F^{\hat{\mathbf{M}}_\tau}, F^{\check{\mathbf{M}}_\tau}) \\
= \ & L^3(F^{\mathbf{AB}^*+\mathbf{BA}^*} - F^{\check{\mathbf{A}}\check{\mathbf{B}}^*+\check{\mathbf{B}}\check{\mathbf{A}}^*}) \\
\le \ & \frac{1}{N}\mathrm{tr}[(\mathbf{AB}^* + \mathbf{BA}^* - (\check{\mathbf{A}}\check{\mathbf{B}}^* + \check{\mathbf{B}}\check{\mathbf{A}}^*))(\mathbf{AB}^* + \mathbf{BA}^* - (\check{\mathbf{A}}\check{\mathbf{B}}^* + \check{\mathbf{B}}\check{\mathbf{A}}^*))^*] \\
\le \ & \frac{2}{N}\mathrm{tr}[(\mathbf{AB}^* - \check{\mathbf{A}}\check{\mathbf{B}}^*)(\mathbf{AB}^* - \check{\mathbf{A}}\check{\mathbf{B}}^*)^*] + \frac{2}{N}\mathrm{tr}[(\mathbf{BA}^* - \check{\mathbf{B}}\check{\mathbf{A}}^*)(\mathbf{BA}^* - \check{\mathbf{B}}\check{\mathbf{A}}^*)^*]
\end{aligned}$$

and

$$\begin{aligned}
& \frac{1}{N}\mathrm{tr}[(\mathbf{AB}^* - \check{\mathbf{A}}\check{\mathbf{B}}^*)(\mathbf{AB}^* - \check{\mathbf{A}}\check{\mathbf{B}}^*)^*] \\
= \ & \frac{1}{N}\mathrm{tr}[((\mathbf{A} - \check{\mathbf{A}})\mathbf{B}^* + \check{\mathbf{A}}(\mathbf{B} - \check{\mathbf{B}})^*)((\mathbf{A} - \check{\mathbf{A}})\mathbf{B}^* + \check{\mathbf{A}}(\mathbf{B} - \check{\mathbf{B}})^*)^*] \\
\le \ & \frac{2}{N}\mathrm{tr}[((\mathbf{A} - \check{\mathbf{A}})\mathbf{B}^*)((\mathbf{A} - \check{\mathbf{A}})\mathbf{B}^*)^* + (\check{\mathbf{A}}(\mathbf{B} - \check{\mathbf{B}})^*)(\check{\mathbf{A}}(\mathbf{B} - \check{\mathbf{B}})^*)^*]
\end{aligned}$$

Define $J_2 = \frac{1}{N}\mathrm{tr}((\mathbf{A} - \check{\mathbf{A}})\mathbf{B}^*)((\mathbf{A} - \check{\mathbf{A}})\mathbf{B}^*)^*$, then we have

$$\begin{aligned}
J_2 \ = \ & \frac{1}{N}\mathrm{tr}((\mathbf{A} - \check{\mathbf{A}})^*(\mathbf{A} - \check{\mathbf{A}}))(\mathbf{B}^*\mathbf{B}) \\
= \ & \frac{C}{T^3}\sum_{i=1}^{N}\sum_{j=1}^{N}|\sum_{k=1}^{T}(\check{\varepsilon}_{ik} - \hat{\varepsilon}_{ik})\bar{\hat{\varepsilon}}_{j(k+\tau)}|^2 \\
= \ & \frac{C}{T^3}\sum_{i=1}^{N}\sum_{j=1}^{N}(\sum_{k_1=1}^{T}(\check{\varepsilon}_{ik_1} - \hat{\varepsilon}_{ik_1})\bar{\hat{\varepsilon}}_{j(k_1+\tau)})(\sum_{k_2=1}^{T}(\bar{\check{\varepsilon}}_{ik} - \bar{\hat{\varepsilon}}_{ik})\hat{\varepsilon}_{j(k_2+\tau)}) \\
= \ & \frac{C}{T^3}\sum_{i=1}^{N}\sum_{j=1}^{N}[\sum_{k=1}^{T}|\check{\varepsilon}_{ik} - \hat{\varepsilon}_{ik}|^2|\hat{\varepsilon}_{j(k+\tau)}|^2 + \\
& \sum_{k_1,k_2=1,k_1>k_2}^{T}(\check{\varepsilon}_{ik_1} - \hat{\varepsilon}_{ik_1})(\bar{\check{\varepsilon}}_{ik_2} - \bar{\hat{\varepsilon}}_{ik_2})\bar{\hat{\varepsilon}}_{j(k_1+\tau)}\hat{\varepsilon}_{j(k_2+\tau)} + \\
& \sum_{k_1,k_2=1,k_1<k_2}^{T}(\check{\varepsilon}_{ik_1} - \hat{\varepsilon}_{ik_1})(\bar{\check{\varepsilon}}_{ik_2} - \bar{\hat{\varepsilon}}_{ik_2})\bar{\hat{\varepsilon}}_{j(k_1+\tau)}\hat{\varepsilon}_{j(k_2+\tau)}] \\
\equiv \ & J_{21} + J_{22} + J_{23}.
\end{aligned}$$

Furthermore,

$$\begin{aligned} J_{21} &= \frac{C}{T^3}\sum_{j=1}^{N}[\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\in E}|\check{\varepsilon}_{ik}-\hat{\varepsilon}_{ik}|^2|\hat{\varepsilon}_{j(k+\tau)}|^2+ \\ &\quad \sum_{i=1,\dots N,k=1,\dots,T,(i,k)\notin E}|\check{\varepsilon}_{ik}-\hat{\varepsilon}_{ik}|^2|\hat{\varepsilon}_{j(k+\tau)}|^2] \\ &\equiv J_{211}+J_{212} \end{aligned}$$

By definition of E, we have $\frac{1-\sigma_{ik}^2}{\Delta}>1$ for any $(i,k)\in E$ and therefore

$$\begin{aligned} \mathrm{E}J_{211} &= \frac{C}{T^3}\sum_{j=1}^{N}\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\in E}\mathrm{E}|\check{\varepsilon}_{ik}-\hat{\varepsilon}_{ik}|^2\mathrm{E}|\hat{\varepsilon}_{j(k+\tau)}|^2 \\ &\le \frac{C}{T^3}\sum_{j=1}^{N}\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\in E}\mathrm{E}|\hat{\varepsilon}_{j(k+\tau)}|^2 \\ &\le \frac{C}{T^3}\sum_{j=1}^{N}\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\in E}\frac{1-\sigma_{ik}^2}{\Delta}\mathrm{E}|\hat{\varepsilon}_{j(k+\tau)}|^2 \\ &\le \frac{C}{T^3}\sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{k=1}^{T}T^{-\frac{\delta}{4+2\delta}} \\ &= O(T^{-\frac{\delta}{4+2\delta}}). \end{aligned}$$

For any $(i,k)\notin E$, we have

$$(1-\sigma_{ik}^{-1})^2=\frac{(\sigma_{ik}-1)^2}{\sigma_{ik}^2}=\frac{(1-\sigma_{ik}^2)^2}{\sigma_{ik}^2(1+\sigma_{ik})^2}\le C(1-\sigma_{ik}^2)^2\le C\eta^{-2\delta}T^{-\frac{2\delta}{2+\delta}},$$

Here and in what follows, we assume that $\eta\to 0$ slow enough such that the above upper bound tends to 0 as $T\to\infty$. This together with $\mathrm{E}\hat{\varepsilon}_{ij}^2<\infty$ implies

$$\begin{aligned} \mathrm{E}J_{212} &= \frac{C}{T^3}\sum_{j=1}^{N}\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\notin E}(1-\sigma_{ik}^{-1})^2|\mathrm{E}\hat{\varepsilon}_{ik}|^2|\mathrm{E}\hat{\varepsilon}_{j(k+\tau)}|^2 \\ &\le \frac{C}{T^3}\sum_{j=1}^{N}\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\notin E}(1-\sigma_{ik}^{-1})^2 \\ &= O(T^{-\frac{2\delta}{2+\delta}}) \end{aligned}$$

Note that summands in J_{22} and J_{23} are pairwise orthogonal, hence we have $\mathrm{E}J_{22} = \mathrm{E}J_{23} = 0$. Therefore, we have $\mathrm{E}J_2 \to 0$.

Now, we want to compute $\mathrm{Var}J_2$. First, we have

$$\begin{aligned}
\mathrm{Var}(J_{211}) &= \frac{C}{T^6}\sum_{j=1}^{N}\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\in E} \mathrm{E}|\check{\varepsilon}_{ik}-\hat{\varepsilon}_{ik}|^4\mathrm{E}|\hat{\varepsilon}_{j(k+\tau)}|^4 \\
&\le \frac{C}{T^6}\sum_{j=1}^{N}\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\in E} (\mathrm{E}|\check{\varepsilon}_{ik}|^4+\mathrm{E}|\hat{\varepsilon}_{ik}|^4)\mathrm{E}|\hat{\varepsilon}_{j(k+\tau)}|^4 \\
&\le \frac{C}{T^6}\sum_{j=1}^{N}\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\in E} \mathrm{E}|\hat{\varepsilon}_{ik}|^4\mathrm{E}|\hat{\varepsilon}_{j(k+\tau)}|^4 \\
&= O(T^{-1-\frac{4\delta}{2+\delta}})
\end{aligned}$$

For simplicity, write

$$\begin{aligned}
J_{212} &= \sum_{j=1}^{N}\sum_{i=1,\dots N,k=1,\dots,T,(i,k)\notin E} (1-\sigma_{ik}^{-1})^2[(|\hat{\varepsilon}_{ik}|^2-\sigma_{ik}^2)(|\hat{\varepsilon}_{j(k+\tau)}|^2-\sigma_{j(k+\tau)}^2)+ \\
&\quad \sigma_{ik}^2(|\hat{\varepsilon}_{j(k+\tau)}|^2-\sigma_{j(k+\tau)}^2)+\sigma_{j(k+\tau)}^2(|\hat{\varepsilon}_{ik}|^2-\sigma_{ik}^2)+\sigma_{ik}^2\sigma_{j(k+\tau)}^2] \\
&\equiv J_{2121}+J_{2122}+J_{2123}+J_{2124}
\end{aligned}$$

$$\begin{aligned}
J_{22} &= \frac{C}{T^3}[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k_1,k_2=1,k_1>k_2,k_1\neq k_2+\tau}^{T} (\check{\varepsilon}_{ik_1}-\hat{\varepsilon}_{ik_1})(\bar{\check{\varepsilon}}_{ik_2}-\bar{\hat{\varepsilon}}_{ik_2})\bar{\hat{\varepsilon}}_{j(k_1+\tau)}\hat{\varepsilon}_{j(k_2+\tau)}+ \\
&\quad \sum_{i,j=1,i\neq j}^{N}\sum_{k_2=1}^{T}(\check{\varepsilon}_{i(k_2+\tau)}-\hat{\varepsilon}_{i(k_2+\tau)})(\bar{\check{\varepsilon}}_{ik_2}-\bar{\hat{\varepsilon}}_{ik_2})\bar{\hat{\varepsilon}}_{j(k_2+2\tau)}\hat{\varepsilon}_{j(k_2+\tau)}+ \\
&\quad \sum_{i=1}^{N}\sum_{k_2=1}^{T}(\check{\varepsilon}_{i(k_2+\tau)}-\hat{\varepsilon}_{i(k_2+\tau)})(\bar{\check{\varepsilon}}_{ik_2}-\bar{\hat{\varepsilon}}_{ik_2})\bar{\hat{\varepsilon}}_{i(k_2+2\tau)}\hat{\varepsilon}_{i(k_2+\tau)} \\
&\equiv J_{221}+J_{222}+J_{223}
\end{aligned}$$

Note that in all expressions except J_{2124}, components are orthogonal to each other. In addition, as a constant, J_{2124} does not contribute to $\mathrm{Var}J_2$.

Therefore, we have

$$\begin{aligned}
\mathrm{Var}J_{2121} &\leq \frac{C}{T^6}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{T}(1-\sigma_{ik}^{-1})^4(\mathrm{E}|\hat{\varepsilon}_{ik}|^4)^2 \\
&\leq \frac{C}{T^6}T^3T^{-\frac{4\delta}{2+\delta}}T^{\frac{2(2-\delta)}{2+\delta}} \\
&= O(T^{-1-\frac{8\delta}{2+\delta}})
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}J_{2122} &= \mathrm{Var}J_{2123} \\
&\leq \frac{C}{T^6}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{T}(1-\sigma_{ik}^{-1})^4\sigma_{ik}^4\mathrm{E}|\hat{\varepsilon}_{ik}|^4 \\
&\leq \frac{C}{T^6}T^3T^{-\frac{4\delta}{2+\delta}}T^{\frac{2-\delta}{2+\delta}} \\
&= O(T^{-2-\frac{6\delta}{2+\delta}})
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}J_{221} &\leq \frac{C}{T^6}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k_1,k_2=1,k_1>k_2,k_1\neq k_2+\tau}^{T}\mathrm{E}|\check{\varepsilon}_{ik_1}-\hat{\varepsilon}_{ik_1}|^2\mathrm{E}|\bar{\check{\varepsilon}}_{ik_2}-\bar{\hat{\varepsilon}}_{ik_2}|^2\mathrm{E}|\bar{\hat{\varepsilon}}_{j(k_1+\tau)}|^2\mathrm{E}|\hat{\varepsilon}_{j(k_2+\tau)}|^2 \\
&\leq \frac{C}{T^6}T^4 \\
&= O(T^{-2}) \\
\mathrm{Var}J_{222} &\leq \frac{C}{T^6}\sum_{i,j=1,i\neq j}^{N}\sum_{k_2=1}^{T}\mathrm{E}|\check{\varepsilon}_{i(k_2+\tau)}-\hat{\varepsilon}_{i(k_2+\tau)}|^2\mathrm{E}|\bar{\check{\varepsilon}}_{ik_2}-\bar{\hat{\varepsilon}}_{ik_2}|^2\mathrm{E}|\bar{\hat{\varepsilon}}_{j(k_2+2\tau)}|^2\mathrm{E}|\hat{\varepsilon}_{j(k_2+\tau)}|^2 \\
&\leq \frac{C}{T^6}T^3 \\
&= O(T^{-3}) \\
\mathrm{Var}J_{223} &\leq \frac{C}{T^6}\sum_{i=1}^{N}\sum_{k_2=1}^{T}\mathrm{E}|(\check{\varepsilon}_{i(k_2+\tau)}-\hat{\varepsilon}_{i(k_2+\tau)})\hat{\varepsilon}_{i(k_2+\tau)}|^2\mathrm{E}|\bar{\check{\varepsilon}}_{ik_2}-\bar{\hat{\varepsilon}}_{ik_2}|^2\mathrm{E}|\bar{\hat{\varepsilon}}_{i(k_2+2\tau)}|^2 \\
&\leq \frac{C}{T^6}\sum_{i=1}^{N}\sum_{k_2=1}^{T}\mathrm{E}|\hat{\varepsilon}_{i(k_2+\tau)}|^4 \\
&\leq \frac{C}{T^6}T^2T^{\frac{2-\delta}{2+\delta}} \\
&= O(T^{-3-\frac{2\delta}{2+\delta}})
\end{aligned}$$

Therefore, we have $\mathrm{Var}J_2 = O(T^{-1-\varepsilon})$ for some $\varepsilon > 0$. Thus, $J_2 \to 0$ a.s.. Similarly, we have

$$\mathrm{E}\frac{1}{N}\mathrm{tr}(\check{\mathbf{A}}(\mathbf{B} - \check{\mathbf{B}})^*)(\check{\mathbf{A}}(\mathbf{B} - \check{\mathbf{B}})^*)^* \to 0$$

and

$$\mathrm{Var}\frac{1}{N}\mathrm{tr}(\check{\mathbf{A}}(\mathbf{B} - \check{\mathbf{B}})^*)(\check{\mathbf{A}}(\mathbf{B} - \check{\mathbf{B}})^*)^* = O(T^{-1-\varepsilon'})$$

for some $\varepsilon' > 0$. Hence, we have $\frac{1}{N}\mathrm{tr}[(\mathbf{A}\mathbf{B}^* - \check{\mathbf{A}}\check{\mathbf{B}}^*)(\mathbf{A}\mathbf{B}^* - \check{\mathbf{A}}\check{\mathbf{B}}^*)^*] \to 0$ a.s.. By interchanging $\mathbf{A}$ and $\mathbf{B}$, we have $\frac{1}{N}\mathrm{tr}[(\mathbf{B}\mathbf{A}^* - \check{\mathbf{B}}\check{\mathbf{A}}^*)(\mathbf{B}\mathbf{A}^* - \check{\mathbf{B}}\check{\mathbf{A}}^*)^*] \to 0$ a.s.. Therefore, $L^3(F^{\hat{\mathbf{M}}_\tau}, F^{\check{\mathbf{M}}_\tau}) \to 0$ a.s..

APPENDIX B: SOME TECHNICAL LEMMAS

LEMMA B.1. *Under the assumptions of Theorem 1.1,* $\boldsymbol{\gamma}_k^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_k - \frac{c}{2}m_n(z) \to 0$ *almost surely and uniformly in* $k \le T + \tau$, *where* $m_n(z) = \frac{1}{N}tr\mathbf{A}^{-1}$.

The proof of Lemma B.1 is similar to the proof of Lemma 9.1 of Bai and Silverstein (2010).

PROOF. Write $\mathbf{A}_k^{-1} = (a_{ij})$. For any given $r \ge 1$, we have

$$\mathrm{E}|\boldsymbol{\gamma}_k^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_k - \frac{1}{2T}\mathrm{tr}\mathbf{A}_k^{-1}|^{2r} \le 2^{2r-1}(\mathrm{E}|S_1|^{2r} + \mathrm{E}|S_2|^{2r}),$$

where $S_1 = \frac{1}{2T}\sum_{i=1}^N a_{ii}(\varepsilon_{ki}^2 - 1)$ and $S_2 = \frac{1}{2T}\sum_{1\le i\ne j\le N} a_{ij}\varepsilon_{ki}\varepsilon_{kj}$.

By noting $|a_{ii}| \le ||\mathbf{A}_k^{-1}|| \le v^{-1}$ and $\mathrm{E}|\varepsilon_{ki}^2 - 1|^{(2+\delta)/2} \equiv M < \infty$, we get

$$\begin{aligned}
\mathrm{E}|S_1|^{2r} &= \frac{1}{4^r T^{2r}} \mathrm{E}(|\sum_{i=1}^{N}(\varepsilon_{ki}^2 - 1)a_{ii}|^{2r}) \\
&\le \frac{1}{4^r T^{2r}} \sum_{l=1}^{r} \sum_{1\le j_1<\cdots<j_l\le N} \sum_{\substack{i_1+\cdots i_l=2r,\\ i_1\ge 2,\cdots,i_l\ge 2}} (2r)! \prod_{t=1}^{l} \frac{\mathrm{E}|\varepsilon_{kj_t}^2 - 1|^{i_t}|a_{j_tj_t}|^{i_t}}{i_t!} \\
&\le \frac{\eta^{4r}}{4^r v^{2r} T^{\frac{2\delta r}{2+\delta}}} \sum_{l=1}^{r} \eta^{-(2+\delta)l} T^{-l} N^l M^l l^{2r}. \qquad \text{(B.1)}
\end{aligned}$$

Next, let us consider

$$\mathrm{E}|S_2|^{2r} = \frac{1}{4^r T^{2r}} \sum a_{i_1j_1}\bar{a}_{t_1\ell_1} \cdots a_{i_rj_r}\bar{a}_{t_r\ell_r} \mathrm{E}(\varepsilon_{ki_1}\varepsilon_{kt_1}\varepsilon_{kj_1}\varepsilon_{k\ell_1} \cdots \varepsilon_{k\ell_r}\varepsilon_{kt_r}\varepsilon_{kj_r}\varepsilon_{k\ell_r}).$$

Draw a directional graph G of $2r$ edges that link i_s to j_s and ℓ_s to t_s, $s = 1, \cdots r$. Note that if G has a vertex whose degree is 1, then the graph corresponds to a term with expectation 0. That is, for any nonzero term, the vertex degrees of the graph are not less than 2. Write the non-coincident vertices as $v_1, \cdots, v_m$ with degrees $p_1, \cdots, p_m$ greater than 1. We have $m \le r$. By assumption, we have

$$|\mathrm{E}(\varepsilon_{ki_1}\varepsilon_{kt_1}\varepsilon_{kj_1}\varepsilon_{k\ell_1} \cdots \varepsilon_{k\ell_r}\varepsilon_{kt_r}\varepsilon_{kj_r}\varepsilon_{k\ell_r})| \le (\eta^2 T^{2/(2+\delta)})^{r-m}.$$

Now, suppose that the graph consists of q connected components $G_1, \cdots, G_q$ with $m_1, \cdots, m_q$ noncoincident vertices, respectively. Let us consider the con- tribution by G_1 to $\mathrm{E}|S_2|^r$. Assume that G_1 has s_1 edges, $e_1, \cdots, e_{s_1}$. Choose a tree G_1' from G_1, and assume its edges are $e_1, , e_{m_1-1}$, without loss of generality. Note that

$$\sum_{v_1,\cdots,v_{m_1}\le N} \prod_{t=1}^{m_1-1} |a_{e_t}|^2 \le ||\mathbf{A}^{-1}||^{2m_1-2} N \le \frac{N}{v^{2m_1-2}}$$

and

$$\sum_{v_1,\cdots,v_{m_1}\le N}\prod_{t=m_1}^{s_1}|a_{e_t}|^2\le\frac{N^{m_1-1}}{v^{2s_1-2m_1+2}}.$$

Here, the first inequality follows from the fact that $\sum_{v_1}|a_{v_1v_2}|^2\le||\mathbf{A}^{-1}||^2\le v^{-2}$ since it is a diagonal element of $\mathbf{A}^{-1}(\mathbf{A}^{-1})^*$. The second inequality follows from the fact that $\sum_{v_1}|a_{v_1v_2}|^\ell\le v^{-\ell}$ for any $\ell\ge 2$ and that $s_1\ge m_1$ since all vertices have degrees not less than 2. Therefore, the contribution of G_1 is bounded by

$$\sum_{v_1,\cdots,v_{m_1}\le N}\prod_{t=1}^{s_1}|a_{e_t}|\le\Big(\sum_{v_1,\cdots,v_{m_1}\le N}\prod_{t=1}^{m_1-1}|a_{e_t}|^2\sum_{v_1,\cdots,v_{m_1}\le N}\prod_{t=m_1}^{s_1}|a_{e_t}|^2\Big)^{1/2}\le\frac{N^{m_1/2}}{v^{s_1}}.$$

Noting that $m_1+\cdots+m_q=m$ and $s_1+\cdots+s_q=2r$, eventually we obtain that the contribution of the isomorphic class for a given canonical graph is $\frac{N^{m/2}}{v^{2r}}$. Because the two vertices of each edge cannot coincide, we have $q\le m/2$. The number of canonical graphs is less than $\binom{m}{2}^{2r}\le m^{4r}$. We finally obtain

$$\begin{aligned}\mathrm{E}|S_2|^{2r} &\le \frac{1}{4^r v^{2r}T^{2r}}\sum_{m=2}^{r}N^{m/2}(\eta^2T^{2/(2+\delta)})^{2r-m}m^{4r}\\ &\le \frac{1}{4^r v^{2r}T^{\frac{2\delta r}{2+\delta}}}\sum_{m=2}^{r}(\frac{N^{1/2}}{\eta^2T^{2/(2+\delta)}})^m m^{4r}.\end{aligned}\tag{B.2}$$

Using (B.1) and (B.2), for any $t>0$, there exists $r>t/\delta+t/2$ such that $\mathrm{E}|\boldsymbol{\gamma}_k^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_k-\frac{1}{2T}\mathrm{tr}\mathbf{A}_k^{-1}|^{2r}=O(T^{-t})$. Therefore, by Borel-Cantelli lemma,

$$\boldsymbol{\gamma}_k^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_k-\frac{1}{2T}\mathrm{tr}\mathbf{A}_k^{-1}\to 0\tag{B.3}$$

almost surely and uniformly in $k\le T+\tau$.

Let F_N denote the ESD of $\mathbf{M}_\tau$ and F_{Nk} the ESD of $\mathbf{M}_\tau-\boldsymbol{\gamma}_k(\boldsymbol{\gamma}_{k+\tau}+\boldsymbol{\gamma}_{k-\tau})^*-(\boldsymbol{\gamma}_{k+\tau}+\boldsymbol{\gamma}_{k-\tau})\boldsymbol{\gamma}_k^*$.

By Theorem A.43 of Bai and Silverstein (2010), we have

$$\|F_N - F_{Nk}\| \leq \frac{4}{N},$$

where $\|f\| = \sup_x |f(x)|$. Thus

$$\begin{aligned} |\frac{1}{N}\big(\mathrm{tr}(\mathbf{A}^{-1}) - \mathrm{tr}(\mathbf{A}_k^{-1})\big)| &= |\int \frac{1}{x-u-iv} d(F_N - F_{NK})| \\ &\leq \frac{1}{v}\|F_N - F_{Nk}\| \leq \frac{4}{vN}. \end{aligned}$$

This implies that

$$\frac{1}{N}\big(\mathrm{tr}(\mathbf{A}^{-1}) - \mathrm{tr}(\mathbf{A}_k^{-1})\big) \to 0,\ a.s.$$

uniformly in $k \leq T + \tau$. Substituting the above into (B.3), the proof of the lemma is complete.

LEMMA B.2. *Under the assumptions of Theorem 1.1, we have* $\boldsymbol{\gamma}_k^* \mathbf{A}_k^{-1} \boldsymbol{\gamma}_l \to 0$, *almost surely and uniformly in* $k \neq l$.

PROOF. Let $\mathbf{A}_k^{-1}\boldsymbol{\gamma}_l = \mathbf{b} = (b_1, \cdots, b_N)^T$. Noting $\varepsilon_{lj} < \eta_N T^{1/(2+\delta)}$ and $\mathrm{E}|\varepsilon_{lj}|^{2+\delta} = v_{2+\delta} < \infty$, we have

$$\begin{aligned} &\mathrm{E}(\boldsymbol{\gamma}_k^* \mathbf{A}_k^{-1} \boldsymbol{\gamma}_l)^{2r} = \frac{1}{2^r T^r} \mathrm{E}(|\sum_{i=1}^N \varepsilon_{ki} b_i|^{2r}) \\ \leq\ & \frac{1}{2^r T^r} \mathrm{E} \sum_{l=1}^r \sum_{1 \leq j_1 < \cdots < j_l \leq N} \sum_{i_1 + \cdots i_l = 2r} \frac{(2r)!}{i_1! \cdots i_l!} |\varepsilon_{kj_1}^{i_1} b_{j_1}^{i_1} \cdots \varepsilon_{kj_l}^{i_l} b_{j_1}^{i_l}| \\ \leq\ & \frac{\eta^{2r}}{2^r T^{\frac{\delta r}{2+\delta}}} \mathrm{E} \sum_{l=1}^r \eta^{-(2+\delta)l} T^{-l} v_{2+\delta}^l \sum_{1 \leq j_1 < \cdots < j_l < N} \sum_{\substack{i_1 + \cdots i_l = 2r, \\ i_1 \geq 2, \cdots, i_l \geq 2}} \frac{(2r)!}{i_1! \cdots i_l!} |b_{j_1}|^{i_1} \cdots |b_{j_l}|^{i_l}. \end{aligned}$$

By $\sum_{j=1}^{N}|b_j|^2=\|\mathbf{A}_k^{-1}\boldsymbol{\gamma}_l\|^2$ and Cauchy-Schwartz inequality, we have

$$\begin{aligned}
&\sum_{1\le j_1<\cdots<j_l\le N}\ \sum_{\substack{i_1+\cdots i_l=2r,\\ i_1\ge 2,\cdots,i_l\ge 2}}\frac{(2r)!}{i_1!\cdots i_l!}|b_{j_1}|^{i_1}\cdots|b_{j_l}|^{i_l}\\
\le&\sum_{\substack{i_1+\cdots i_l=2r,\\ i_1\ge 2,\cdots,i_l\ge 2}}\frac{(2r)!}{i_1!\cdots i_l!}(\sum_{j=1}^{N}|b_j|^2)^r\\
\le&\ l^{2r}||\mathbf{A}_k^{-1}||^{2r}||\boldsymbol{\gamma}_l||^{2r}\\
\le&\ \frac{l^{2r}}{v^{2r}}||\boldsymbol{\gamma}_l||^{2r}.
\end{aligned}$$

Noting $\varepsilon_{lj}<\eta_N T^{1/(2+\delta)}$ and $\mathrm{E}|\varepsilon_{lj}|^{2+\delta}=v_{2+\delta}<\infty$, we get

$$\begin{aligned}
\mathrm{E}||\boldsymbol{\gamma}_l||^{2r} &= \frac{1}{2^rT^r}\mathrm{E}(\sum_{j=1}^{N}\varepsilon_{lj}^2)^r\\
&= \mathrm{E}\frac{1}{2^rT^r}\sum_{l=1}^{r}\sum_{1\le j_1<\cdots<j_l\le N}\ \sum_{i_1+\cdots i_l=r}\frac{r!}{i_1!\cdots i_l!}\varepsilon_{lj_1}^{2i_1}\cdots\varepsilon_{lj_l}^{2i_l}\\
&\le \mathrm{E}\frac{1}{2^r}\sum_{l=1}^{r}v_{2+\delta}^l\eta^{2r-(2+\delta)l}T^{-\frac{\delta r}{2+\delta}-l}\sum_{1\le j_1<\cdots<j_l\le N}\ \sum_{i_1+\cdots i_l=r}\frac{r!}{i_1!\cdots i_l!}\\
&\le \frac{\eta^{2r}}{2^rT^{\frac{\delta r}{2+\delta}}}\sum_{l=1}^{r}(\frac{\eta^{2+\delta}T}{Nv_{2+\delta}})^{-l}l^r.
\end{aligned}$$

For any $t>0$, there exists $r>2t/\delta+t$ such that $\mathrm{E}|\boldsymbol{\gamma}_k^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_l|^r=O(T^{-t})$. Therefore by Borel-Cantelli lemma,

$$\boldsymbol{\gamma}_k^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_l\to 0$$

almost surely and uniformly in $k\neq l$. The proof of the lemma is complete.

In the next lemma, we find the limit of $\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_{k+\tau}$ when $T-k\to\infty$ and that of $\boldsymbol{\gamma}_{k-\tau}^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_{k-\tau}$ when $k\to\infty$.

Lemma B.3. *Assume that* $\frac{1}{N}tr(\mathbf{A}-z\mathbf{I})^{-1}\to m=m(z)$. *When* $T-k\to\infty$,

$$\boldsymbol{\gamma}_{k+\tau}^*\mathbf{A}_k^{-1}\boldsymbol{\gamma}_{k+\tau}\to\frac{\frac{c}{2}m}{x_1},$$

where x_1 *is the root of the quadratic equation* $x^2 - x + \frac{1}{4}c^2m^2 = 0$ *with the larger absolute value.*

When $k \to \infty$,

$$\boldsymbol{\gamma}_{k-\tau}^* \mathbf{A}_k^{-1} \boldsymbol{\gamma}_{k-\tau} \to \frac{\frac{c}{2}m}{x_1},$$

where x_1 *is the same as above.*

PROOF. Write $a = \frac{c}{2}m$, $W_k = \boldsymbol{\gamma}_{k+\tau}^* \mathbf{A}_k^{-1} \boldsymbol{\gamma}_{k+\tau}$ and $W_{k,k+\tau,\cdots,k+\ell\tau} = \boldsymbol{\gamma}_{k+(\ell+1)\tau}^* \mathbf{A}_{k,k+\tau,\cdots,k+\ell\tau}^{-1} \boldsymbol{\gamma}_{k+(\ell+1)\tau}$. Then by (3.4), we have

$$W_k = \frac{a + r(k)}{1 - aW_{k,k+\tau}},$$

where $r(k) = o_{a.s.}(1)$, uniformly in $k \le T + \tau$. Using this relation again, we obtain

$$\text{(B.4)} \quad W_k = \frac{a + r(k)}{1 - a\frac{a + r(k+\tau)}{1 - aW_{k,k+\tau,k+2\tau}}} = \frac{(a + r(k))(1 - aW_{k,k+\tau,k+2\tau})}{1 - aW_{k,k+\tau,k+2\tau} - a(a + r(k+\tau))}.$$

Applying this relation ℓ times, we may express W_k in the following form

$$W_k = \frac{(a + r(k))(\alpha_{k,\ell-1} - a\alpha_{k,\ell-2}W_{k,k+\tau,\cdots,k+\ell\tau})}{\alpha_{k,\ell} - a\alpha_{k,\ell-1}W_{k,k+\tau,\cdots,k+\ell\tau}}$$

where the coefficients satisfy the recursive relation

$$\text{(B.5)} \quad \alpha_{k,\ell} = \alpha_{k,\ell-1} - a(a + r(k+\ell\tau))\alpha_{k,\ell-2}, \ \alpha_{k,1} = 1, \ \alpha_{k,0} = 1.$$

Define x_1 and x_0 as the roots of the equation $x^2 = x - a^2$ with $|x_1| > |x_0|$(Note that the equal sign happens only when $a = \pm\frac{1}{2}$ which is impossible for $\Im(z) > 0$ because a is the Stieltjes transform of a distribution function). Then we have

$$x_1 = \begin{cases} \frac{1}{2}(1 - \sqrt{1 - 4a^2}) & if \Im(a^2) > 0, \\ \frac{1}{2}(1 + \sqrt{1 - 4a^2}) & if \Im(a^2) < 0, \end{cases}, \quad x_0 = \begin{cases} \frac{1}{2}(1 + \sqrt{1 - 4a^2}) & if \Im(a^2) > 0, \\ \frac{1}{2}(1 - \sqrt{1 - 4a^2}) & if \Im(a^2) < 0. \end{cases}$$

Similarly, define $\nu_{k,1}$ and $\nu_{k,0}$ as the the roots of the equation $x^2 = x - a(a+r(k))$, with $|\nu_{k,1}| > |\nu_{k,0}|$. By this definition, we have

$$\nu_{k,1} = \begin{cases} \frac{1}{2}\big(1 - \sqrt{1-4a(a+r(k))}\big) & if \Im(a(a+r(k))\big) > 0, \\ \frac{1}{2}\big(1 + \sqrt{1-4a(a+r(k))}\big) & if \Im(a(a+r(k))) < 0. \end{cases}$$

Further, define α such that $\alpha\nu_{k,1} + (1-\alpha)\nu_{k,0} = 1$. Then, define $\nu_{k,1,j} = \nu_{k,j}$, for $j = 0, 1$. For $t \geq 1$, define

$$\nu_{k,t+1,j} = 1 - \frac{a(a+r(k+(t+1)\tau)}{\nu_{k,t,j}}, \text{ for } j = 0, 1. \tag{B.6}$$

From this, we have

$$\alpha\nu_{k,1,1} + (1-\alpha)\nu_{k,1,0} = 1 = \alpha_{k,1}.$$

And

$$\begin{aligned} & \alpha\nu_{k,2,1}\nu_{k,1,1} + (1-\alpha)\nu_{k,2,0}\nu_{k,1,0} \\ = \; & \alpha(\nu_{k,1,1} - a(a+r(k+2\tau))) + (1-\alpha)(\nu_{k,1,0} - a(a+r(k+2\tau))) \\ = \; & \alpha_{k,1} - a(a+r(k+2\tau)) = \alpha_{k,2}. \end{aligned}$$

Using (B.5) and (B.6), we may prove by induction that

$$\alpha\prod_{t=1}^{\ell}\nu_{k,t,1} + (1-\alpha)\prod_{t=1}^{\ell}\nu_{k,t,0} = \alpha_{k,\ell}.$$

Our next goal is to estimate the difference between $\nu_{k,t+1,j}$ and x_j. First, by noticing the definition of $\nu_{k,j}$, we have

$$\nu_{k,j} - x_j = \nu_{k,1,j} - x_j = o_{a.s.}(1)$$

where, and in what follows, the remainder term $o_{a.s.}(1)$ is uniform in k and ℓ. Then, by $\nu_{k,1,j} = \nu_{k,1,j}^2 + a(a + r(k))$ we have

$$\begin{aligned}\nu_{k,2,j} - x_j &= \frac{\nu_{k,1,j} - a(a + r(k + 2\tau))}{\nu_{k,1,j}} - x_j \\ &= \frac{a(r(k) - r(k + 2\tau))}{\nu_{k,1,j}} + \nu_{k,1,j} - x_j = o_{a.s.}(1).\end{aligned}$$

By induction, we can prove that

$$\nu_{k,t+1,j} - x_j = o_{a.s.}(1),$$

provided that t is bounded by a fixed amount M. Therefore, for any given $\eta > 0$, when N is large, we have

$$\prod_{t=1}^{M} \left| \frac{\nu_{k,t+1,0}}{\nu_{k,t+1,1}} \right| \leq \left(\frac{|x_0|}{|x_1|} + \eta \right)^M .$$

Note that $|x_0| < |x_1|$. Thus, for any given $\varepsilon > 0$, we may choose $\eta > 0$ and $\ell < \infty$, such that

$$\prod_{t=1}^{\ell} \left| \frac{\nu_{k,t+1,0}}{\nu_{k,t+1,1}} \right| \leq \varepsilon.$$

That means, when $\ell \to \infty$ slowly, we have $\alpha_{k,\ell+1} = x_1 \alpha_{k,\ell}(1 + o_{a.s.}(1))$. Consequently,

$$\begin{aligned}W_k &= \frac{(a + r(k))(\alpha_{k,\ell} - a\alpha_{k,\ell-1} W_{k,k+\tau,\cdots,k+\ell\tau})}{\alpha_{k,\ell+1} - a\alpha_{k,\ell} W_{k,k+\tau,\cdots,k+\ell\tau}} \\ &= \frac{a}{x_1}(1 + o_{a.s.}(1)).\end{aligned}$$

The first conclusion of the lemma is proved. By duality, the second conclusion follows.

Remark B.1. *Note that one of the key steps in the above proof is to let $\ell \to \infty$ slowly. This may not be possible when $\tau \to \infty$, as in this case we* *may have $\ell\tau > T-k$ and $\boldsymbol{\gamma}_{k+\ell\tau}$ does not exist. However, such ℓ exists when $\tau = o(T)$. Therefore, the proof is still valid in this case.*

LEMMA B.4. *Assume the conditions of Lemma B.3 hold. For all $k \in [1, T+\tau]$, we have*

$$\boldsymbol{\gamma}_{k-\tau}^{*}\mathbf{A}_k^{-1}\boldsymbol{\gamma}_{k+\tau} \to 0, \quad a.s.$$

where the convergence is uniform in k.

PROOF. Obviously, when $\tau < k \le 2\tau$, the lemma is true because $\boldsymbol{\gamma}_{k-\tau}$ is independent of $\mathbf{A}_k$. Similarly, the lemma is true when $T-\tau < k \le T$.

When $2\tau < k \le T-\tau$, by (3.6) and what is proved in the last lemma,

$$\begin{aligned}\boldsymbol{\gamma}_{k-\tau}\mathbf{A}_k^{-1}\boldsymbol{\gamma}_{k+\tau} &= \frac{-a}{1-\frac{a^2}{x_1}+o_{a.s.}(1)}\boldsymbol{\gamma}_{k-\tau}\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau} \\ &= \frac{-a}{x_1+o_{a.s.}(1)}\boldsymbol{\gamma}_{k-\tau}\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}\end{aligned}$$

For $x_1x_0 = a^2$ and $|x_1| > |x_0|$, we have $|a| < |x_1|$. Thus

$$|\boldsymbol{\gamma}_{k-\tau}\mathbf{A}_k^{-1}\boldsymbol{\gamma}_{k+\tau}| \le (\frac{|a|}{|x_1|}+\eta)|\boldsymbol{\gamma}_{k-\tau}\mathbf{A}_{k,k+\tau}^{-1}\boldsymbol{\gamma}_{k+2\tau}|$$

for some $\eta > 0$ such that $|a/x_1| + \eta < 1$. Now, the lemma can be proved by induction.

REFERENCES

[1] Bai, J. and Ng, S. (2002) Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.

[2] Bai, Z. D., Miao, B. Q., and Jin, B. S. (2007) On limit theorem for the eigenvalues of product of two random matrices. *J. Multivariate Anal.* 98(1), 76–101.

[3] Bai, Z. D. and Zhou, W. (2008) Large sample covariance matrices without independence structures in columns. *Statistica Sinica* 18, 425–42.

[4] Bai, Z. D. , Silverstein, J. W. (2004) CLT for linear spectral statistics of large dimensional sample covariance matrices. *Ann. Probab.* 32, 553 – 605.

[5] Bai, Z. D. , Silverstein, J. W. (2010) Spectral Analysis of Large Dimensional Random Matrices, 2nd ed.. *Springer Verlag, New York.*

[6] Bai, Z. D. , Silverstein, J. W. (2012) No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Information-Plus-Noise Type Matrices. *Random Matrices: Theory and Applications* 1, 1150004.

[7] Bai, Z. D. and Yao, J. F. (2008) Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. H. Poincare Probab. Statist.* Volume 44, Number 3, 447-474.

[8] 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.

[9] Baik, Jinho and Silverstein, J. W. (2006) Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* 97(6), 1382-1408.

[10] Bose, A. Mitra, J. (2002) Limiting spectral distribution of a special circulant. *Statistics and Probability Letters* 60, 111–20.

[11] Dozier, R. B. and Silverstein, J. W. (2007a) On the Empirical Distribution of Eigenvalues of Large Dimensional Information-Plus-Noise Type Matrices. *J. Multivariate Anal.* 98(4), 678 – 694.

[12] Dozier, R. B. and Silverstein, J. W. (2007b) Analysis of the Limiting Spectral Distribution of Large Dimensional Information-Plus-Noise Type Matrices. *J. Multivariate Anal.* 98(6), 1099 – 1122.

[13] Forni M. and Lippi, M. (2001) The generalized dynamic factor model. *Econometric Theory* 17(6), 1113–1141 .

[14] Grenander, U. and Silverstein, J. W. (1977) Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.* 32(2), 499–519.

[15] Harding, M. C. (2012) Estimating the number of factors in large dimensional factor models. *Under Review.*

[16] Jiang, T. F. (2004) The limiting distributions of eigenvalues of sample covariance matrices. *Sankhya* 66, 35–48.

[17] Jin, B. S., Wang, C., Miao, B. Q., and Huang, Lo. M. N. (2009) Limiting spectral distribution of large-dimensional sample covariance matrices generated by VARMA. *J. Multivariate Anal.* 100, 2112–25.

[18] Johnstone, I. (2001) On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* 29, 295 – 327.

[19] Jonsson, D. (1982) Some limit theorems for the eigenvalues of a sample covariance matrix. J. Multivariate Anal. 12, 1–38.

[20] Marčhenko, V. A. and Pastur, L. A. (1967) Distribution of eigenvalues for some sets of matrices. *Mat. Sb.* 72, 507 – 536.

[21] Onatski, A. (2009) A formal statistical test for the number of factors in the approximate factor models. *Econometrica* 77(5). 1447-1479.

[22] Onatski, A. (2012) Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics* 168. 244-258.

[23] Raj Rao, N. and Edelman, A. (2008) Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Transactions of Signal Processing* Volume 56, Issue 7. 2625-38.

[24] Silverstein, J. W. (1995) Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* 5, 331–339.

[25] Silverstein, J. W. Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.* 54, 175–192.

[26] Tulino, A. M. Verdu, S. (2004) Random matrix theory and wireless communications. *Jnl. of Comm. and Inf. Theory.* Now Publishers Volume 1, Number 1. 1-182.

[27] Wachter, K. W. (1978) The strong limits of random matrix spectra for sample matrices of independent elements. *Ann Probab.* 6(1), 1–18.

[28] Wachter, K. W. (1980) The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.* 8, 937–957.

[29] Wang, C., Jin, B. S., and Miao, B. Q. (2011) On limiting spectral distribution of large sample covariance matrices by VARMA(p,q), *J. of Time Series Analysis* 32(5), 539–546.

[30] Wigner, E. P. (1955) Characteristic vectors bordered matrices with infinite dimensions. *Ann. Math.* 62, 548– 564.

[31] Wigner, E. P.(1958) On the distributions of the roots of certain symmetric matrices. *Ann. Math.* 67, 325– 327.

[32] Yin, Y. Q. and Krishnaiah, P. R. (1983) A limit theorem for the eigenvalues of product of two random matrices. *J. Multivariate Anal.* 13, 489–507.

[33] Vautard, R., Tiou, P., and Ghil, M. (1992) Singular-spectrum analysis: A toolkit for short, noisy chaotic signal. *Physica D.* Volume 58. 95-126.

[34] Zheng, S. R. (2012) Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. H. Poincare Probab. Statist.* Volume 48, Number 2. 444-476.

[35] Zhigljavsky, A. (2012) Singular spectrum analysis: Present, past and future. *Proc. Intl. Conf. on Forecasting Economic and Financial Systems,* Beijing.

Department of Statistics and Finance
University of Science and
Technology of China.
E-mail: jbs@ustc.edu.cn

Department of Statistics and Applied
Probability
National University of Singapore.
E-mail: a0030755@nus.edu.sg

KLASMOE and School of Math. and Stat.,
Northeast Normal University.
E-mail: baizd@nenu.edu.cn

Department of Civil and
Environmental Engineering,
439 Panama Mall,
Stanford University,
CA-94305
E-mail: kknair@stanford.edu

Department of Economics,
Stanford University,
579 Serra Mall,
Stanford, CA 94305.
Phone: (650) 723-4116;
Fax: (650) 725-5702;
E-mail: mch@stanford.edu
URL: http://www.nenu.edu.cn/professor/pro/show.php?flag=1&id=123