On singular value distribution of large-dimensional autocovariance matrices

Zeng Li*, **Guangming Pan**† **and Jianfeng Yao**‡

*Department of Statistics and Actuarial Science**,‡

The University of Hong Kong

e-mail: u3001205@hku.hk; jeffyao@hku.hk

School of Physical & Mathematical Sciences†

Nanyang Technological University

e-mail: gmpan@ntu.edu.sg

Abstract: Let $(\varepsilon_j)_{j\geq 0}$ be a sequence of independent $p-$dimensional random vectors and $\tau \geq 1$ a given integer. From a sample $\varepsilon_1, \cdots, \varepsilon_{T+\tau}$ of the sequence, the so-called lag-τ auto-covariance matrix is $C_\tau = T^{-1}\sum_{j=1}^{T} \varepsilon_{\tau+j}\varepsilon_j^t$. When the dimension p is large compared to the sample size T, this paper establishes the limit of the singular value distribution of C_τ assuming that p and T grow to infinity proportionally and the sequence has uniformly bounded fourth order moments. Compared to existing asymptotic results on sample covariance matrices developed in random matrix theory, the case of an auto-covariance matrix is much more involved due to the fact that the summands are dependent and the matrix C_τ is not symmetric. Several new techniques are introduced for the derivation of the main theorem.

MSC 2010 subject classifications: Primary 60F99; secondary 62M10, 62H99.

Keywords and phrases: random matrix theory, large-dimensional auto-covariance matrix, limiting spectral distribution, singular value distribution.

1. Introduction

Let $\varepsilon_1, \ldots, \varepsilon_{T+\tau}$ be a sample from a stationary process with values in $\mathbb{R}^p$. The $p \times p$ matrix

$$C_\tau := \frac{1}{T}\sum_{j=1}^{T} \varepsilon_{\tau+j}\varepsilon_j^t, \tag{1.1}$$

is the so-called lag$-\tau$ *sample auto-covariance matrix* of the process (here u^t denotes the transpose of a vector or matrix u). In a classical low-dimensional situation where the dimension p is assumed much smaller than the sample size T, C_τ is very close to $\mathbb{E}\, C_\tau = \mathbb{E}\, \varepsilon_{1+\tau}\varepsilon_1^t$ so that its asymptotic behavior when $T \to \infty$ (so p is considered as fixed) is well known. In the high-dimensional context where typically the dimension p is of same order as T, C_τ will not converge to $\mathbb{E}\, C_\tau$ and

*Corresponding author.

†

‡Research partly supported by GRFGrant7054

its asymptotic properties have not been well investigated. In this paper, we study the empirical spectral distribution (ESD) of C_τ, namely, the distribution generated by its p singular values. The main result of the paper is the establishment of the limit of this ESD when (ε_j) is an independent sequence with elements having a finite fourth moments while p and T grow to infinity proportionally.

In order to understand the importance of the limiting spectral distribution (LSD) of singular values of the auto-covariance matrix C_τ, we describe a statistical problem where these distributions are of central interest. A recent paper Lam and Yao [8] considers the following factor model

$$x_i = \Lambda f_i + \varepsilon_i + \mu, \tag{1.2}$$

where $\{x_i;\ 0 \le i \le T\}$ is an observed p-dimensional sequence, $\{f_i\}$ a sequence of m-dimensional "latent factors" ($m \ll p$) uncorrelated with the error process $\{\varepsilon_i\}$ and $\mu \in \mathbb{R}^p$ is the general mean. A particularly important question here is the determination of the number m of factors. For any stationary process $\{w_i\}$, let $\Sigma_w = \mathrm{cov}(w_i, w_{i-1})$ be its (population) lag-1 auto-covariance matrix. We have

$$\Sigma_x = \Lambda \Sigma_f \Lambda^t.$$

It turns out that Σ_x has exactly m non-null singular values so that based on a sample $x_0, x_1, \ldots, x_T$ it seems natural to infer m from the singular values of the sample lag-1 auto-covariance matrix

$$\begin{aligned}\Gamma_x &= \frac{1}{T}\sum_{j=1}^{T}(\Lambda f_j + \varepsilon_j)(\Lambda f_{j-1} + \varepsilon_{j-1})^t \\ &= \Lambda\left(\frac{1}{T}\sum_{j=1}^{T} f_j f_{j-1}^t\right)\Lambda^t + \Lambda\left(\frac{1}{T}\sum_{j=1}^{T} f_j \varepsilon_{j-1}^t\right) + \left(\frac{1}{T}\sum_{j=1}^{T} \varepsilon_j f_{j-1}^t\right)\Lambda^t + C_1 \ .\end{aligned}$$

Because Λ has rank m, the first three terms all have rank bounded by m and Γ_x appears as a finite-rank perturbation of the lag-1 auto-covariance matrix C_1 which in general has rank $p \gg m$. Therefore, understanding the properties of the singular values of C_1 will be of primary importance for the understanding of the m largest singular values of the matrix of Γ_x which are, as said above, fundamental for the determination of the number of factors m. Actually in a following paper in Li, Wang and Yao [9], we established a phase transition phenomenon: a factor singular value l_i of Γ_x will tend to a limit outside the support of the LSD of C_1 if and only if the corresponding population factor strength exceeds some critical value. Based on this transition phenomenon, we proposed a consistent estimator of the number of factors by counting the number of eigenvalues lying outside that support. Notice however that this statistical problem is given here to describe a potential application of the theory established in this paper, but this theory on singular value distribution is general and can be applied to fields other than statistics.

If we take $\tau = 0$ in (1.1), the matrix $S = \frac{1}{T}\sum_{j=1}^{T} \varepsilon_j \varepsilon_j^t$ is the sample covariance matrix from the observations. The theory for eigenvalue distributions of S has been extensively studied in the random matrix literature dating back to the seminal paper [11] where the famous Marčenko-Pastur law as limit of eigenvalue distributions has been found for a wide class of sample covariance matrices. Further development includes the almost sure convergence of these distributions ([13]) and conditions for convergence of the largest and the smallest eigenvalues. Meanwhile, book-length analysis of sample covariance matrices can be found in [3], [1], [12] and [5]. The situation of an auto-covariance matrix C_τ is however completely different. We know only four references treating auto-covariance matrices, [6], [4],[14] and [10]. All the references considered the LSD of the symmetrized auto-covariance matrix $B = \frac{1}{2}(C_\tau + C_\tau^t)$. The former three assumed that the vectors $\varepsilon_1, \cdots, \varepsilon_{T+\tau}$ are independent, while the latter allowed them to be temporally dependent. It is noticed that the singular value of C_τ are not directly comparable to the eigenvalues of the symmetric part B. Indeed, let $A = \frac{1}{2}(C_\tau - C_\tau^t)$ be the anti-symmetric part of C_τ. Then $C_\tau C_\tau^t = B^2 - A^2$ and we see that the square of the singular value of C_τ and the square of the eigenvalue of B are different precisely because C_τ is not symmetric, that is $A \neq 0$.

Technically, there are basically two major differences between C_τ and S. First, while S is a non-negative symmetric random matrix, C_τ is even not symmetric and we must rely on singular value distributions which are in general much more involved than eigenvalue distributions. Secondly, because of the positive lag τ, the summands in C_τ are no more independent as it is the case for the sample covariance matrix S. This again makes the analysis of C_τ more difficult. As a consequence of these major differences, several new techniques are introduced in the paper in order to complete the proofs, although the general strategy is common in the random matrix theory (see Bai and Silverstein [3], Pastur and Shcherbina [12]). For example, the characterization of the Stieltjes transform of the limiting distribution is obtained via a system of equations due to the time delay τ where for the case of sample covariance matrix, the characterization is given by a single equation([11], [13]).

The rest of the paper is organized as follows. The main theorem of the paper is introduced in Section 2. Section 3 details the proof of the main theorem when time lag $\tau = 1$. Section 4 generalizes the proof from time lag $\tau = 1$ to any given positive number. Meanwhile, in contrast to other aspects discussed above, the preliminary steps of truncation, centralization and standardization of the matrix entries are similar to the case of a sample covariance matrix. They are given in Appendix A. To ease the reading of the proofs, technical lemmas are grouped in Section 5.

2. Main Results

In this paper, we intend to derive the limiting singular value distribution of the lag$-\tau$ auto-covariance matrix defined in (1.1). It will be done in two steps. We derive the main result first for the lag-1($\tau = 1$) sample auto-covariance matrix $C_1 = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_j \varepsilon_{j-1}^t$. It turns out that the general case $\tau \geq 1$ is essentially the same and the extension is easily obtained. The details of the extension are given in Section 4.

Therefore, we consider the lag-1 sample auto-covariance matrix $C_1 = \frac{1}{T}\sum_{j=1}^{T}\varepsilon_j\varepsilon_{j-1}^t$. By definition, it is equivalent to study the limiting spectral distribution (LSD) of the matrix

$$A = C_1C_1^t = \frac{1}{T^2}(\sum_{j=1}^{T}\varepsilon_j\varepsilon_{j-1}^t)(\sum_{j=1}^{T}\varepsilon_{j-1}\varepsilon_j^t).$$

Alternatively,

$$A = \frac{1}{T^2}XY^tYX^t,$$

where $X = (\varepsilon_1, \cdots, \varepsilon_T)_{p\times T}$, $Y = (\varepsilon_0, \cdots, \varepsilon_{T-1})_{p\times T}$. Here we define a modified version of the A matrix,

$$B = \frac{1}{T^2}Y^tYX^tX = \sum_{j=1}^{p}s_js_j^t\sum_{j=1}^{p}r_jr_j^t,$$

where $s_j = \frac{1}{\sqrt{T}}\left(\varepsilon_{j0}, \varepsilon_{j1}, \cdots, \varepsilon_{j,T-1}\right)^t$ is the j-th row of Y, and $r_j = \frac{1}{\sqrt{T}}\left(\varepsilon_{j1}, \varepsilon_{j2}, \cdots, \varepsilon_{j,T}\right)^t$ the j-th row of X. As A and B have same nonzero eigenvalues, the LSD of A can be derived from the LSD of B.

The main result of the paper is

Theorem 2.1. *Let the following assumptions hold:*

(a) $\varepsilon_i = \left(\varepsilon_{1i}, \cdots \varepsilon_{pi}\right)^t, i = 0, 1, 2, \cdots, T$ *are independent p-dimensional real-valued random vectors with independent entries satisfying condition:*

$$\mathbb{E}(\varepsilon_{it}) = 0,\ \mathbb{E}(\varepsilon_{it}^2) = 1, \quad \sup_{1\leq i\leq p, 0\leq t\leq T}\mathbb{E}\left(|\varepsilon_{it}|^{4+\delta}\right) < M,$$

for some constant M and arbitrarily small positive δ;

(b) *As $p \to \infty$, the sample size $T = T(p) \to \infty$ and $p/T \to c > 0$.*

Then,

(1) *as $p, T \to \infty$, almost surely, the empirical spectral distribution F^B of B, converges to a non-random probability distribution $\underline{F}$ whose Stieltjes transform $x = x(\alpha)$, $\alpha \in \mathbb{C}\setminus\mathbb{R}$, satisfies the equation*

$$\alpha^2x^3 - 2\alpha\left(c-1\right)x^2 + \left(c-1\right)^2x - \alpha x - 1 = 0. \tag{2.1}$$

(2) *Moreover, for $\alpha \in \mathbb{C}^+ = \{z : \mathfrak{Im} z > 0\}$, equation (2.1) admits a unique solution $\alpha \mapsto x(\alpha)$ with positive imaginary part and the density function of the LSD $\underline{F}$ is:*

$$f(u) = \frac{1}{\pi u}\left\{-u - \frac{5(c-1)^2}{3} + \frac{2^{4/3}(3u+(c-1)^2)(c-1)}{3d(u)^{1/3}} + \frac{2^{2/3}(c-1)d(u)^{1/3}}{3} + \frac{1}{48}\left[-8(c-1) + \frac{2\times 2^{1/3}(3u+(c-1)^2)}{d(u)^{1/3}} + 2^{2/3}d(u)^{1/3}\right]^2\right\}^{1/2},$$

where $d(u) = -2(c-1)^3 + 9(1+2c)u + 3\sqrt{3}\sqrt{u(-4u^2+(-1+4c(5+2c))u - 4c(c-1)^3)}$.

Moreover, the support of f(u) is $(0,b]$ *for* $0 < c \le 1$, *and* $[a,b]$ *for* $c > 1$, *where*

$$a = \frac{1}{8}(-1+20c+8c^2-(1+8c)^{3/2}), \quad b = \frac{1}{8}(-1+20c+8c^2+(1+8c)^{3/2}).$$

It's easy to check that when $c < 1$, the LSD of B has a point mass $1-c$ at the origin since $\text{rank}(B) = p < T$ for large p and T, and at the same time we have

$$\begin{cases} \int_0^b f(u)du = c, & 0 < c < 1, \\ \int_a^b f(u)du = 1, & c \ge 1. \end{cases}$$

Since the matrix A we are interested in has the same non-zero eigenvalues as B, the following proposition holds.

Proposition 2.1. *Under the conditions of Theorem 2.1, the ESD of A converges a.s. to a non-random limit distribution*

$$F = \frac{1}{c}\underline{F} + (1-\frac{1}{c})\delta_0,$$

whose Stieltjes transform $y = y(\alpha)$, $\alpha \in \mathbb{C}\setminus\mathbb{R}$, *satisfies the equation*

$$\alpha^2c^2y^3 + \alpha c(c-1)y^2 - \alpha y - 1 = 0.$$

In particular, F *has the density function*

$$\begin{cases} \frac{1}{c}f(u), \ u \in (0,b], \ \textit{for } 0 < c < 1, \\ \frac{1}{c}f(u), \ u \in [a,b], \ \textit{for } c \ge 1. \end{cases}$$

where in the later case $c \ge 1$, F *has an additional mass* $(1-\frac{1}{c})$ *at the origin.*

The following details the density function of LSD of A for different values of c.

- When $c = 1$, the support is $0 \le u \le \frac{27}{4}$ and the density function is

$$\frac{1}{c}f(u) = \frac{1}{\pi u}\left[-u + 3\left(\frac{u}{2^{2/3}d(u)^{1/3}} + \frac{d(u)^{1/3}}{6\times 2^{1/3}}\right)^2\right]^{1/2},$$

  where $d(u) = 27u + 3\sqrt{3}\times\sqrt{u(-4u^2+27u)}$. It's easy to see that as $u \to 0_+$, $f(u) \to \infty$.
- If $c < 1$, it can be seen from the explicit form of $f(u)$ that when $u \to 0_+$, $\frac{1}{c}f(u) \to \infty$ because the u in the denominator of the density function cannot be completely canceled out.

- If $c > 1$, the shape of the density function turns out to be a little different from the case $c \leq 1$. Nevertheless it's quite intuitive because the lower bound of the support is positive and the density function is bounded.

The density functions of LSD of A for $c = 0.5, 1, 2, 3$ are displayed on Figure 1.

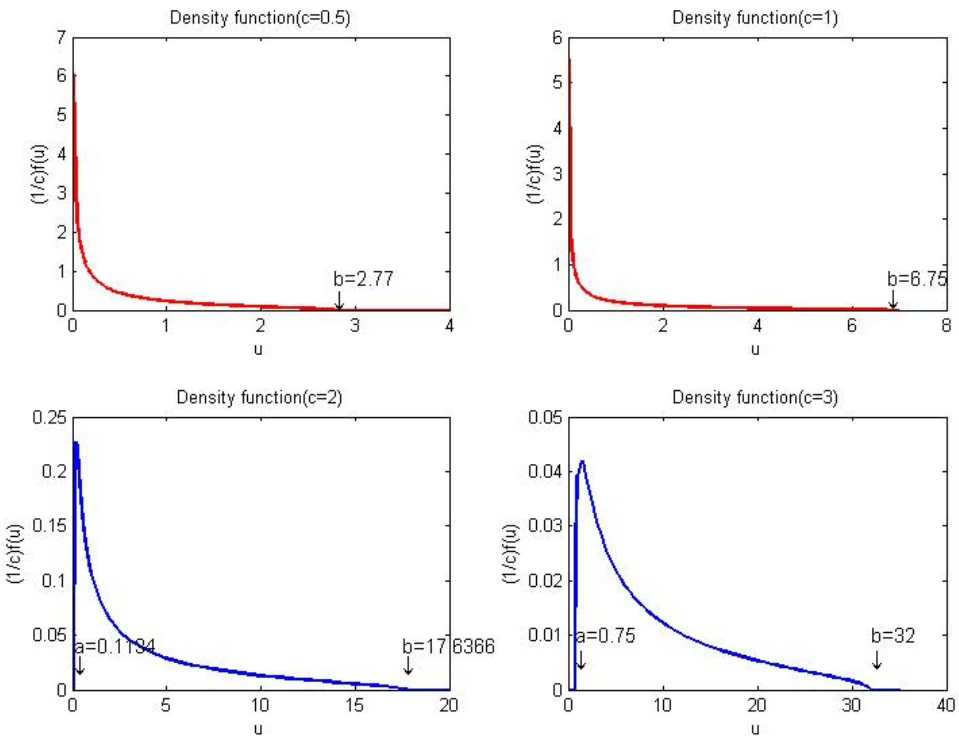

FIG 1. *Density plots of the LSD of B.Top to bottom and left to right: c=0.5,1,2 and 3, respectively*

3. Proofs

3.1. Proof of Theorem 2.1

The proof of the theorem follows the general strategy based on the Stieltjes transform as presented in Silverstein [13], Bai and Silverstein [3] and Pastur and Shcherbina [12]. However, the random matrix B here is no longer a covariance matrix as considered in these references. Almost all the steps of the proof need new arguments and ideas compared to the case of sample covariance matrices considered so far in the literature. Following this method, the first step is to truncate the entries $\{\varepsilon_{jt}\}$ at a convenient rate using Assumption (a). After truncation and the follow-up steps of centralization and standardization, we may assume that, for some constant M, η and arbitrarily small δ,

$$|\varepsilon_{ij}| \le \eta T^{1/4}, \quad \mathbb{E}(\varepsilon_{ij}) = 0, \quad Var(\varepsilon_{ij}) = 1, \quad \sup_{1\le i\le p, 0\le j\le T} \mathbb{E}\left(|\varepsilon_{ij}|^{4+\delta}\right) < M.$$

The details of these technical steps are given in Appendix A.

By the rank inequality (Theorem A.44 of [3]), it is enough to consider

$$B = \sum_{j=1}^{p} s_j s_j^t \sum_{j=1}^{p} r_j r_j^t = P_1 \tilde{C} P_1^t \tilde{C},$$

where

$$s_j = P_1 r_j = \frac{1}{\sqrt{T}}(0, \varepsilon_{j1}, \cdots, \varepsilon_{j,T-1})^t, \quad C = \sum_{j=1}^{p} s_j s_j^t, \quad \tilde{C} = \sum_{j=1}^{p} r_j r_j^t, \quad P_1 = \begin{pmatrix} \mathbf{0} & 0 \\ \mathbf{I}_{T-1} & \mathbf{0} \end{pmatrix}.$$

At this stage, the important observation is that here we have replaced $s_j = \frac{1}{\sqrt{T}}(\varepsilon_{j0}, \varepsilon_{j1}, \cdots, \varepsilon_{j,T-1})^t$ by $\tilde{s}_j = \frac{1}{\sqrt{T}}(0, \varepsilon_{j1}, \cdots, \varepsilon_{j,T-1})^t$ without altering the LSD of B since when $T \to \infty$, the effect of this substitution will vanish. For the sake of convenience, we still use s_j to denote $\tilde{s}_j$.

For $\alpha \in \mathbb{C} \setminus \mathbb{R}$, define

$$B(\alpha) = \sum_{j=1}^{p} s_j s_j^t \sum_{j=1}^{p} r_j r_j^t - \alpha \mathbf{I}_T.$$

Let

$$x_0 = \frac{1}{T} tr(B^{-1}(\alpha)), \quad y_0 = \frac{1}{T} tr(\tilde{C} B^{-1}(\alpha)), \quad z_0 = \frac{1}{T} tr(B^{-1}(\alpha) C).$$

The method consists in finding a system of two asymptotic equations satisfied by x_0 and y_0. Solving the system yields an asymptotic equivalent for x_0 and finally leads to the equation (2.1) satisfied by the limit of x_0. Meanwhile, x_0 is the Stieltjes transform of the matrix B which can be recovered from the inversion formula.

Let

$$B_j(\alpha) = \sum_{k\neq j} s_k s_k^t \sum_{i\neq j} r_i r_i^t - \alpha \mathbf{I}_T, \quad C_j = C - s_j s_j^t, \quad \tilde{C}_j = \tilde{C} - r_j r_j^t, \quad 1 \le j \le p,$$

then

$$\begin{aligned} B(\alpha) &= B_j(\alpha) + \sum_{i\neq j} s_j s_j^t r_i r_i^t + \sum_{k\neq j} s_k s_k^t r_j r_j^t + s_j s_j^t r_j r_j^t \\ &= B_j(\alpha) + s_j s_j^t \tilde{C}_j + C_j r_j r_j^t + s_j s_j^t r_j r_j^t. \end{aligned}$$

We have

$$\mathbf{I}_T = B(\alpha) B^{-1}(\alpha) = \left(\sum_{j=1}^{p} s_j s_j^t\right) \left(\sum_{j=1}^{p} r_j r_j^t\right) B^{-1}(\alpha) - \alpha B^{-1}(\alpha).$$

Taking the trace and dividing both sides by T, we get

$$1=\frac{1}{T}\sum_{j=1}^{p}s_j^t\tilde{C}B^{-1}(\alpha)s_j-\alpha\frac{1}{T}tr\left(B^{-1}(\alpha)\right). \tag{3.1}$$

Note that $x_0=\frac{1}{T}tr(B^{-1}(\alpha))$ is the Stieltjes transform of the ESD of the matrix B, and its limit will be found by letting $p,T\to\infty$ on both sides of the equation.

Consider $s_j^t\tilde{C}B^{-1}(\alpha)s_j$, using the identities

$$\left(B+\sum_{j=1}^{m}ab_j^t\right)^{-1}a=\frac{B^{-1}a}{1+\sum_{j=1}^{m}b_j^tB^{-1}a},$$

we have

$$\begin{aligned}s_j^t\tilde{C}B^{-1}(\alpha)s_j&=\frac{s_j^t\tilde{C}\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}s_j}{1+s_j^t\tilde{C}\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}s_j}\\&=1-\frac{1}{1+s_j^t\tilde{C}_j\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}s_j+s_j^tr_jr_j^t\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}s_j}\\&:=1-\frac{1}{1+L_1+L_2},\end{aligned}$$

where L_1 and L_2 are implicitly defined.

For L_1, by the following equation

$$B^{-1}-D^{-1}=B^{-1}(D-B)D^{-1},$$

and Lemma 5.1, or equivalently by Lemma 2.7 of [2], we have

$$\begin{aligned}L_1&=s_j^t\tilde{C}_j\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}s_j\\&=s_j^t\tilde{C}_jB_j^{-1}(\alpha)s_j-s_j^t\tilde{C}_jB_j(\alpha)^{-1}C_jr_jr_j^t\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}s_j\\&=s_j^t\tilde{C}_jB_j^{-1}(\alpha)s_j-\frac{s_j^t\tilde{C}_jB_j^{-1}(\alpha)C_jr_jr_j^tB_j(\alpha)^{-1}s_j}{1+r_j^tB_j^{-1}(\alpha)C_jr_j}\\&=\frac{1}{T}tr\left(\tilde{C}_jB_j^{-1}(\alpha)\right)-\frac{\frac{1}{T}tr\left(\tilde{C}_jB_j^{-1}(\alpha)C_jP_1^t\right)\cdot\frac{1}{T}tr\left(B_j^{-1}(\alpha)P_1\right)}{1+\frac{1}{T}tr\left(B_j(\alpha)^{-1}C_j\right)}+o_{a.s.}(1).\end{aligned}$$

For L_2, we have

$$\begin{aligned}L_2&=s_j^tr_jr_j^t\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}s_j=s_j^tr_jr_j^tB_j^{-1}(\alpha)s_j-\frac{s_j^tr_jr_j^tB_j^{-1}(\alpha)C_jr_jr_j^tB_j^{-1}(\alpha)s_j}{1+r_j^tB_j^{-1}(\alpha)C_jr_j}\\&=\left(r_j^tP_1^tr_j\right)\cdot\frac{1}{T}tr\left(B_j^{-1}(\alpha)P_1\right)-\frac{\left(r_j^tP_1^tr_j\right)\cdot\frac{1}{T}tr\left(B_j^{-1}(\alpha)C_j\right)\cdot\frac{1}{T}tr\left(B_j^{-1}(\alpha)P_1\right)}{1+\frac{1}{T}tr\left(B_j^{-1}(\alpha)C_j\right)}+o_{a.s.}(1)=o_{a.s.}(1).\end{aligned}$$

Therefore, by equation (3.1), we have

$$\begin{aligned}&1+\alpha\frac{1}{T}tr(B^{-1}(\alpha))=o_{a.s.}(1)+ \qquad (3.2)\\&\frac{p}{T}\left(1-\frac{1+\frac{1}{T}tr(B^{-1}(\alpha)C)}{\left(1+\frac{1}{T}tr\left(B^{-1}(\alpha)C\right)\right)\left(1+\frac{1}{T}tr(\tilde{C}B^{-1}(\alpha))\right)-\frac{1}{T}tr\left(\tilde{C}B^{-1}(\alpha)CP_1^t\right)\cdot\frac{1}{T}tr\left(B^{-1}(\alpha)P_1\right)}\right)\end{aligned}$$

Here, we have used the following equivalences, uniformly in j, as $p, T \to \infty$,

$$\frac{1}{T}tr\left(B_j^{-1}(\alpha)C_j\right) = z_0 + o_{a.s.}(1),$$

$$\frac{1}{T}tr\left(B_j^{-1}(\alpha)\right) = x_0 + o_{a.s.}(1),$$

$$\frac{1}{T}tr\left(\tilde{C}_j B_j^{-1}(\alpha)\right) = y_0 + o_{a.s.}(1).$$

Similar to equation (3.1), we have

$$1 = \frac{1}{T}\sum_{j=1}^{p} r_j^t B^{-1}(\alpha) C r_j - \alpha\frac{1}{T}tr\left(B^{-1}(\alpha)\right). \tag{3.3}$$

Considering $r_j^t B^{-1}(\alpha) C r_j$, we have

$$\begin{aligned} r_j^t B^{-1}(\alpha) C r_j &= \frac{r_j^t\left(B_j(\alpha) + s_j s_j^t \tilde{C}_j\right)^{-1} C r_j}{1 + r_j^t\left(B_j(\alpha) + s_j s_j^t \tilde{C}_j\right)^{-1} C r_j} \\ &= 1 - \frac{1}{1 + r_j^t\left(B_j(\alpha) + s_j s_j^t \tilde{C}_j\right)^{-1} C_j r_j + r_j^t\left(B_j(\alpha) + s_j s_j^t \tilde{C}_j\right)^{-1} s_j s_j^t r_j} \\ &:= 1 - \frac{1}{1 + W_1 + W_2}, \end{aligned}$$

where W_1 and W_2 are implicitly defined.

For W_1, we have

$$\begin{aligned} W_1 &= r_j^t\left(B_j(\alpha) + s_j s_j^t \tilde{C}_j\right)^{-1} C_j r_j \\ &= r_j^t B_j^{-1}(\alpha) C_j r_j - r_j^t B_j^{-1}(\alpha) s_j s_j^t \tilde{C}_j\left(B_j(\alpha) + s_j s_j^t \tilde{C}_j\right)^{-1} C_j r_j \\ &= r_j^t B_j^{-1}(\alpha) C_j r_j - \frac{r_j^t B_j^{-1}(\alpha) s_j s_j^t \tilde{C}_j B_j^{-1}(\alpha) C_j r_j}{1 + s_j^t \tilde{C}_j B_j^{-1}(\alpha) s_j} \\ &= \frac{1}{T}tr\left(B_j^{-1}(\alpha)C_j\right) - \frac{\frac{1}{T}tr\left(\tilde{C}_j B_j^{-1}(\alpha) C_j P_1^t\right)\cdot\frac{1}{T}tr\left(B_j^{-1}(\alpha)P_1\right)}{1 + \frac{1}{T}tr\left(\tilde{C}_j B_j(\alpha)^{-1}\right)} + o_{a.s.}(1). \end{aligned}$$

For W_2, we have

$$\begin{aligned} W_2 &= r_j^t\left(B_j(\alpha) + s_j s_j^t \tilde{C}_j\right)^{-1} s_j s_j^t r_j = r_j^t B_j^{-1}(\alpha) s_j s_j^t r_j - \frac{r_j^t B_j^{-1}(\alpha) s_j s_j^t \tilde{C}_j B_j^{-1}(\alpha) s_j s_j^t r_j}{1 + s_j^t \tilde{C}_j B_j^{-1}(\alpha) s_j} \\ &= \left(s_j^t P_1^t s_j\right)\cdot\frac{1}{T}tr\left(B_j^{-1}(\alpha)P_1\right) - \frac{\left(s_j^t P_1^t s_j\right)\cdot\frac{1}{T}tr\left(\tilde{C}_j B_j^{-1}(\alpha)\right)\cdot\frac{1}{T}tr\left(B_j^{-1}(\alpha)P_1\right)}{1 + \frac{1}{T}tr\left(\tilde{C}_j B_j^{-1}(\alpha)\right)} + o_{a.s.}(1) = o_{a.s.}(1). \end{aligned}$$

Therefore, by equation (3.3), we have

$$1 + \alpha\frac{1}{T}tr(B^{-1}(\alpha)) = o_{a.s.}(1) + \tag{3.4}$$
$$\frac{p}{T}\left(1 - \frac{1 + \frac{1}{T}tr(B^{-1}(\alpha)\tilde{C})}{\left(1 + \frac{1}{T}tr\left(B^{-1}(\alpha)C\right)\right)\left(1 + \frac{1}{T}tr(\tilde{C}B^{-1}(\alpha))\right) - \frac{1}{T}tr\left(\tilde{C}B^{-1}(\alpha)CP_1^t\right)\cdot\frac{1}{T}tr\left(B^{-1}(\alpha)P_1\right)}\right)$$

Thus, according to equations (3.2) and (3.4), we have

$$\frac{1}{T}tr(B^{-1}(\alpha)\tilde{C}) = \frac{1}{T}tr(B^{-1}(\alpha)C) + o_{a.s.}(1).$$

By Lemma 5.2, the second term of L_1, $\frac{\frac{1}{T}tr\left(\tilde{C}_j B_j^{-1}(\alpha)C_j P_1^t\right)\cdot\frac{1}{T}tr\left(B_j^{-1}(\alpha)P_1\right)}{1+\frac{1}{T}tr\left(B_j(\alpha)^{-1}C_j\right)}$ is $o_{a.s.}(1)$ since both $\frac{1}{T}tr\left(P_1^t\tilde{C}_j B_j(\alpha)^{-1}C_j\right)$ and $\frac{1}{T}tr\left(B_j(\alpha)^{-1}C_j\right)$ are non-negative and bounded as $p,T\to\infty$, therefore,

$$L_1 = \frac{1}{T}tr\left(\tilde{C}_j B_j^{-1}(\alpha)\right) + o_{a.s.}(1) = y_0 + o_{a.s.}(1).$$

Finally, by equation (3.2), we find

$$1+\alpha x_0 = \frac{p}{T}\left(1-\frac{1}{1+y_0}\right) + o_{a.s.}(1). \tag{3.5}$$

To find a second equation satisfied by x_0 and y_0, using Lemma 5.1 and Lemma 5.2,

$$\begin{aligned}
\frac{1}{T}tr(\tilde{C}B^{-1}(\alpha)) &= \frac{1}{T}tr(\sum_{j=1}^{p} r_j r_j^t B^{-1}(\alpha)) = \frac{1}{T}\sum_{j=1}^{p} r_j^t B^{-1}(\alpha) r_j \\
&= \frac{1}{T}\sum_{j=1}^{p}\frac{r_j^t\left(B_j(\alpha)+s_j s_j^t\tilde{C}_j\right)^{-1}r_j}{1+r_j^t\left(B_j(\alpha)+s_j s_j^t\tilde{C}_j\right)^{-1}C_j r_j + r_j^t\left(B_j(\alpha)+s_j s_j^t\tilde{C}_j\right)^{-1}s_j s_j^t r_j} \\
&= \frac{1}{T}\sum_{j=1}^{p}\frac{r_j^t B_j^{-1}(\alpha) r_j - \dfrac{r_j^t B_j^{-1}(\alpha)s_j s_j^t\tilde{C}_j B_j^{-1}(\alpha)r_j}{1+s_j^t\tilde{C}_j B_j^{-1}(\alpha)s_j}}{1+r_j^t B_j^{-1}(\alpha)C_j r_j - \dfrac{r_j^t B_j^{-1}(\alpha)s_j s_j^t\tilde{C}_j B_j^{-1}(\alpha)C_j r_j}{1+s_j^t\tilde{C}_j B_j^{-1}(\alpha)s_j}} + o_{a.s.}(1) \\
&= \frac{p}{T}\cdot\frac{\dfrac{1}{T}tr(B^{-1}(\alpha))}{1+\dfrac{1}{T}tr(B^{-1}(\alpha)C)} + o_{a.s.}(1).
\end{aligned}$$

This leads to

$$y_0 = \frac{p}{T}\cdot\frac{x_0}{1+y_0} + o_{a.s.}(1). \tag{3.6}$$

In conclusion, (x_0,y_0) satisfy the system

$$\begin{cases} 1+\alpha x_0 = \dfrac{c y_0}{1+y_0} + o_{a.s.}(1), \\ y_0 = \dfrac{c x_0}{1+y_0} + o_{a.s.}(1). \end{cases}$$

Notice that for any T, $|x_0| \le \frac{1}{|\Im\mathfrak{m}(\alpha)|}$ is bounded, and by equation (3.6), $|y_0|$ is also bounded as $T\to\infty$, otherwise (3.6) cannot hold. Therefore, both $\{x_0\}$ and $\{y_0\}$ are bounded sequences. Let be two subsequences $\{x_{t_n}\},\{y_{t_n}\}$ so that $x_{t_n}\to x$ and $y_{t_n}\to y$ as $n\to\infty$. It can be concluded that the limiting functions (x,y) satisfy the system of equations:

$$\begin{cases} 1+\alpha x = \dfrac{c y}{1+y} & (1) \\ y = \dfrac{c x}{1+y} & (2) \end{cases}$$

By eliminating y, we finally find the equation (2.1) satisfied by the limiting function x. Denote by $\mathcal{F}$ all the analytic functions $\{f:\ \mathbb{C}^+ \mapsto \mathbb{C}^+\}$. Because according to the following proof we have one unique solution on $\mathcal{F}$ that satisfies equation (2.1), the whole bounded sequence $\{x_0\}$ has one unique limit x in $\mathcal{F}$.

As for the second statement of Theorem 2.1, in order to find the density function of the LSD $\underline{\mathrm{F}}$ of B, we use the inversion formula:

$$f(u) = \lim_{\varepsilon \to 0_+} \frac{1}{\pi} \mathfrak{Im} x(u + i\varepsilon)$$

where $x(\cdot)$ is the Stieltjes transform of $\underline{\mathrm{F}}$. Write

$$\lim_{\varepsilon \to 0_+} x(u + i\varepsilon) = \psi(u) + i\phi(u),$$

both ψ and ϕ are real-valued functions of u. By substituting $\alpha = u + i\varepsilon$, $x = \psi + i\phi$ into equation (2.1) and letting $\varepsilon \to 0_+$, both the real part and the imaginary part of the LHS of equation (2.1) should equal to 0, i.e.

$$\begin{cases} u^2\psi^3 - 3u^2\psi \cdot \phi^2 - 2u(c-1)\left(\psi^2 - \phi^2\right) - \left(u - (c-1)^2\right)\psi - 1 = 0 & (3) \\ -u^2\phi^2 + 3u^2\psi^2 - 4u(c-1)\psi - \left(u - (c-1)^2\right) = 0 & (4) \end{cases}$$

By plugging (4) into (3), we get

$$-8u^2\psi^3 + 16u(c-1)\psi^2 + (2u - 10(c-1)^2)\psi + \frac{2(c-1)^3}{u} - 2c + 1 = 0.$$

Solving this equation and substituting for ψ in (4), we get

$$\phi_1^2(u) = \frac{1}{u^2}\left\{-u - \frac{5(c-1)^2}{3} + \frac{2^{4/3}(3u + (c-1)^2)(c-1)}{3d(u)^{1/3}} + \frac{2^{2/3}(c-1)d(u)^{1/3}}{3} + \frac{1}{48}\left[-8(c-1) + \frac{2 \times 2^{1/3}(3u + (c-1)^2)}{d(u)^{1/3}} + 2^{2/3}d(u)^{1/3}\right]^2\right\},$$

$$\phi_2^2(u) = \frac{1}{u^2}\left\{-u - \frac{5(c-1)^2}{3} + \frac{1+i\sqrt{3}}{2} \cdot \frac{2^{4/3}(3u + (c-1)^2)(c-1)}{3d(u)^{1/3}} + \frac{1-i\sqrt{3}}{2} \cdot \frac{2^{2/3}(c-1)d(u)^{1/3}}{3} + \frac{1}{48}\left[-8(c-1) + \frac{1+i\sqrt{3}}{2} \cdot \frac{2 \times 2^{1/3}(3u + (c-1)^2)}{d(u)^{1/3}} + \frac{1-i\sqrt{3}}{2} \cdot 2^{2/3}d(u)^{1/3}\right]^2\right\},$$

$$\phi_3^2(u) = \frac{1}{u^2}\left\{-u - \frac{5(c-1)^2}{3} + \frac{1-i\sqrt{3}}{2} \cdot \frac{2^{4/3}(3u + (c-1)^2)(c-1)}{3d(u)^{1/3}} + \frac{1+i\sqrt{3}}{2} \cdot \frac{2^{2/3}(c-1)d(u)^{1/3}}{3} + \frac{1}{48}\left[-8(c-1) + \frac{1-i\sqrt{3}}{2} \cdot \frac{2 \times 2^{1/3}(3u + (c-1)^2)}{d(u)^{1/3}} + \frac{1+i\sqrt{3}}{2} \cdot 2^{2/3}d(u)^{1/3}\right]^2\right\},$$

where

$$d(u) = -2(c-1)^3 + 9(1+2c)u + 3\sqrt{3}\sqrt{u(-4u^2 + (-1 + 4c(5+2c))u - 4c(c-1)^3)}. \qquad (3.7)$$

It can be checked that only the first solution is compatible with the fact that both ψ and ϕ are real-valued functions of u, i.e.

$$\phi^2(u) = \frac{1}{u^2}\left\{-u - \frac{5(c-1)^2}{3} + \frac{2^{4/3}(3u+(c-1)^2)(c-1)}{3d(u)^{1/3}} + \frac{2^{2/3}(c-1)d(u)^{1/3}}{3} + \frac{1}{48}\left[-8(c-1) + \frac{2\times 2^{1/3}(3u+(c-1)^2)}{d(u)^{1/3}} + 2^{2/3}d(u)^{1/3}\right]^2\right\}.$$

From the explicit form of $\phi^2(u)$ we see that, necessarily,

$$u(-4u^2 + (-1+4c(5+2c))u - 4c(c-1)^3) \geq 0,$$

since $u \geq 0$. Solving this quadratic inequality, we get two roots,

$$a = \frac{1}{8}(-1+20c+8c^2-(1+8c)^{3/2}), \quad b = \frac{1}{8}(-1+20c+8c^2+(1+8c)^{3/2}). \tag{3.8}$$

It's very easy to see that a is an increasing function of c and $a = 0$ when $c = 1$.

In other words, if $0 < c < 1$, $-\frac{1}{4} < a < 0$, then the support of the density function should be $(0, b)$. If $c \geq 1$, $a \geq 0$, then the support of the density function is (a, b).

Then the density function of the limiting spectral distribution of the $T \times T$ dimensional multiplied lag-1 sample auto-covariance matrix B is

$$f(u) = \frac{1}{\pi u}\left\{-u - \frac{5(c-1)^2}{3} + \frac{2^{4/3}(3u+(c-1)^2)(c-1)}{3d(u)^{1/3}} + \frac{2^{2/3}(c-1)d(u)^{1/3}}{3} + \frac{1}{48}\left[-8(c-1) + \frac{2\times 2^{1/3}(3u+(c-1)^2)}{d(u)^{1/3}} + 2^{2/3}d(u)^{1/3}\right]^2\right\}^{1/2},$$

where $0 < u \leq b$, for $0 < c \leq 1$ and $a \leq u \leq b$, for $c > 1$, with (a, b) given in equation (3.7) and $d(u)$ given in equation (3.8). Therefore, equation (2.1) admits at least one solution $\alpha \mapsto x(\alpha)$ that corresponds to this density function of the LSD $\underline{F}$. As for the uniqueness, suppose there exists another solution $x_1(\alpha)$ that satisfies equation (2.1), then there should be another density $f_1(u)$ that corresponds to $x_1(\alpha)$ while $f_1(u) \neq f(u)$. However, it can be seen from the previous deductions that the density function is unique. Therefore, $f_1(u) = f(u)$, $x_1(\alpha) = x(\alpha)$. Equation (2.1) admits one unique solution.

3.2. Proof of Proposition 2.1

Under the same conditions in **Theorem 2.1**, the ESD of A converges to a non-random limit distribution F with Stieltjes transform $y = y(\alpha)$. On the other hand, the ESD of B converges to $\underline{F}$ with Stieltjes transform $x = x(\alpha)$ satisfying

$$\alpha^2 x^3 - 2\alpha(c-1)x^2 + (c-1)^2 x - \alpha x - 1 = 0.$$

Since it's known that

$$F = \frac{1}{c}\underline{F} + (1 - \frac{1}{c})\delta_0,$$

conclusively we have

$$(1-c)(-\frac{1}{\alpha}) + cy(\alpha) = x(\alpha).$$

Substituting into the equation of x we can get the equation of y, which is

$$\alpha^2 c^2 y^3 + \alpha c(c-1)y^2 - \alpha y - 1 = 0.$$

4. Extension to lag-τ sample auto-covariance matrix

So far in previous sections, we have focused on the singular value distribution of the lag-1 sample auto-covariance matrix $C_1 = T^{-1}\sum_{j=1}^{T}\varepsilon_j\varepsilon_{j-1}^t$, while in this section, for any given positive integer τ, we discuss the singular value distribution of the lag-τ sample auto-covariance matrix $C_\tau = T^{-1}\sum_{j=1}^{T}\varepsilon_j\varepsilon_{j-\tau}^t$.

Here we follow exactly the same strategy used in the derivation of the LSD of the lag-1 sample auto-covariance matrix. It's easy to see that the difference between C_1 and C_τ lies in that we have now for C_τ,

$$s_j = P_1^\tau r_j = \frac{1}{\sqrt{T}}(\underbrace{0,\cdots,0,}_{\tau\ 0's}\varepsilon_{j1},\cdots,\varepsilon_{j,T-\tau}),\ B = \sum_{j=1}^{p} s_j s_j^t \sum_{j=1}^{p} r_j r_j^t = P_1^\tau \tilde{C}(P_1^\tau)^t\tilde{C}.$$

Meanwhile, the other matrices and notations remain the same using however the new definition of the $s_j's$ above. Consequently, equation (3.2) becomes

$$1 + \alpha\frac{1}{T}tr(B^{-1}(\alpha)) = o_{a.s.}(1) + \tag{4.1}$$
$$\frac{p}{T}\left(1 - \frac{1+\frac{1}{T}tr(B^{-1}(\alpha)C)}{\left(1+\frac{1}{T}tr\left(B^{-1}(\alpha)C\right)\right)\left(1+\frac{1}{T}tr(\tilde{C}B^{-1}(\alpha))\right) - \frac{1}{T}tr\left(\tilde{C}B^{-1}(\alpha)\,C\,(P_1^\tau)^t\right)\cdot\frac{1}{T}tr\left(B^{-1}(\alpha)\,P_1^\tau\right)}\right)$$

Equation (3.4) becomes

$$1 + \alpha\frac{1}{T}tr(B^{-1}(\alpha)) = o_{a.s.}(1) + \tag{4.2}$$
$$\frac{p}{T}\left(1 - \frac{1+\frac{1}{T}tr(B^{-1}(\alpha)\tilde{C})}{\left(1+\frac{1}{T}tr\left(B^{-1}(\alpha)C\right)\right)\left(1+\frac{1}{T}tr(\tilde{C}B^{-1}(\alpha))\right) - \frac{1}{T}tr\left(\tilde{C}B^{-1}(\alpha)\,C\,(P_1^\tau)^t\right)\cdot\frac{1}{T}tr\left(B^{-1}(\alpha)\,P_1^\tau\right)}\right)$$

Thus, according to equation (4.1) and (4.2), we still have

$$\frac{1}{T}tr(B^{-1}(\alpha)\tilde{C}) = \frac{1}{T}tr(B^{-1}(\alpha)C) + o_{a.s.}(1).$$

Meanwhile, by Lemma 5.3, we still have

$$\frac{1}{T} tr\left(B^{-1}(\alpha) P_1^{\tau}\right) = o_{a.s.}(1), \tag{4.3}$$

then by equation (4.1), we have

$$1 + \alpha x_0 = \frac{p}{T}\left(1 - \frac{1}{1+y_0}\right) + o_{a.s.}(1). \tag{4.4}$$

Similarly, as for the second equation satisfied by x_0 and y_0, equation (3.6) persists.

$$y_0 = \frac{p}{T} \cdot \frac{x_0}{1+y_0} + o_{a.s.}(1). \tag{4.5}$$

Therefore, the system of equations satisfied by x_0 and y_0 remains the same when the time lag changes from 1 to τ. In other words, for a given positive time lag τ, the singular value distribution of C_τ is the same with that of C_1 established in Theorem 2.1.

5. TECHNICAL LEMMAS

Lemma 5.1. *Under the same assumptions in **Theorem 2.1**, we have, for any fixed k, $1 \le k < T$, $\forall 1 \le j \le p$, almost surely,*

$$s_j^t B_j^{-1}(\alpha) s_j = \frac{1}{T} tr(B_j^{-1}(\alpha)) + o_{a.s.}(1), \tag{5.1}$$

$$r_j^t B_j^{-1}(\alpha) P_1^k r_j = \frac{1}{T} tr(B_j^{-1}(\alpha) P_1^k) + o_{a.s.}(1), \tag{5.2}$$

$$r_j^t \tilde{C}_j B_j^{-1}(\alpha) P_1^k r_j = \frac{1}{T} tr(\tilde{C}_j B_j^{-1}(\alpha) P_1^k) + o_{a.s.}(1), \tag{5.3}$$

$$s_j^t B_j^{-1}(\alpha) C_j s_j = \frac{1}{T} tr(B_j^{-1}(\alpha) C_j) + o_{a.s.}(1), \tag{5.4}$$

where the $o_{a.s.}(1)$ terms are uniform in $1 \le j \le p$.

Proof. We detail the proof of (5.1) and the proofs of (5.2), (5.3) and (5.4) are very similar, thus omitted.

Denote $B_j^{-1}(\alpha)$ by $(y_{kl}) = Y$, $s_j = \frac{1}{\sqrt{T}}(\varepsilon_{j0}, \cdots, \varepsilon_{j,T-1})^t$, then we have

$$|y_{kl}| < \frac{1}{\nu}, \quad |\varepsilon_{it}| < \eta T^{\frac{1}{4}}, \qquad \sup_{1 \le i \le p, 0 \le t \le T} \mathbb{E}|\varepsilon_{it}|^{4+\delta} < M,$$

where ν is the imaginary part of α.

Following the scheme of Lemma 9.1 of [3] it's easy to see that

$$\begin{aligned}
\mathbb{E}\left|s_j^t Y s_j - \frac{1}{T} tr(Y)\right|^{2r} &= \mathbb{E}\left|\frac{1}{T}\sum_{k,l=1}^{T} \varepsilon_{j,k-1} y_{kl} \varepsilon_{j,l-1} - \frac{1}{T}\sum_{k=1}^{T} y_{kk}\right|^{2r} \\
&= \mathbb{E}\left|\frac{1}{T}\sum_{k=1}^{T}(\varepsilon_{j,k-1}^2 - 1) y_{kk} + \frac{1}{T}\sum_{k \ne l} \varepsilon_{j,k-1} y_{kl} \varepsilon_{j,l-1}\right|^{2r} \\
&= \mathbb{E}\left|S_1 + S_2\right|^{2r} \le 2^r \frac{\mathbb{E}|S_1|^{2r} + \mathbb{E}|S_2|^{2r}}{2},
\end{aligned}$$

where

$$S_1 = \frac{1}{T}\sum_{k=1}^{T}(\varepsilon_{j,k-1}^2 - 1)y_{kk}, \quad S_2 = \frac{1}{T}\sum_{1\le k\ne l\le T} y_{kl}\varepsilon_{j,k-1}\varepsilon_{j,l-1},$$

What's more,

$$\begin{aligned}
\mathbb{E}|S_1|^{2r} &= \mathbb{E}\left|\frac{1}{T}\sum_{k=1}^{T}(\varepsilon_{j,k-1}^2 - 1)y_{kk}\right|^{2r} \\
&\le \frac{1}{T^{2r}}\sum_{t=1}^{r}\sum_{1\le k_1<\cdots<k_t\le T}\sum_{\substack{i_1+\cdots+i_t=2r\\ i_1\ge 2,\cdots,i_t\ge 2}}(2r)!\prod_{l=1}^{t}\frac{\mathbb{E}(\varepsilon_{j,k_l-1}^2-1)^{i_l}y_{k_lk_l}^{i_l}}{i_l!} \\
&\le \frac{1}{T^{2r}}\cdot\frac{1}{v^{2r}}\sum_{t=1}^{r}T^t\sum_{\substack{i_1+\cdots+i_t=2r\\ i_1\ge 2,\cdots,i_t\ge 2}}\frac{(2r)!}{\prod_{l=1}^{t}i_l!}\cdot M^t\frac{(\eta T^{\frac14})^{4r}}{(\eta T^{\frac14})^{4t}} \\
&\le \frac{1}{T^{2r}}\cdot\frac{1}{v^{2r}}\sum_{t=1}^{r}T^t t^{2r}M^t\frac{(\eta T^{\frac14})^{4r}}{(\eta T^{\frac14})^{4t}} = O(\frac{1}{T^r}),
\end{aligned}$$

$$\mathbb{E}|S_2|^{2r} = \frac{1}{T^{2r}}\sum y_{i_1j_1}y_{t_1l_1}\cdots y_{i_rj_r}y_{t_rl_r}\mathbb{E}(\varepsilon_{j,i_1}\varepsilon_{j,j_1}\varepsilon_{j,t_1}\varepsilon_{j,l_1}\cdots\varepsilon_{j,i_r}\varepsilon_{j,j_r}\varepsilon_{j,t_r}\varepsilon_{j,l_r}).$$

Consider a graph G with $2r$ edges that link i_t to j_t and l_t to k_t, $t = 1, \cdots, r$. It's easy to see that 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, then, similarly in Lemma 9.1 of Bai and Silverstein [3], we have,

$$|\mathbb{E}(\varepsilon_{j,i_1}\varepsilon_{j,j_1}\varepsilon_{j,t_1}\varepsilon_{j,l_1}\cdots\varepsilon_{j,i_r}\varepsilon_{j,j_r}\varepsilon_{j,t_r}\varepsilon_{j,l_r})| \le (\eta T^{\frac14})^{2(2r-m)},$$

$$\mathbb{E}|S_2|^{2r} \le \frac{1}{T^{2r}\nu^{2r}}\sum_{m=2}^{r}T^{m/2}(\eta T^{\frac14})^{2(2r-m)}m^{4r} = O(\frac{1}{T^r}).$$

Therefore, by the Borel-Cantelli lemma, we have, $\forall 1 \le j \le p$,

$$s_j^t B_j(\alpha)^{-1}s_j = \frac{1}{T}tr(B_j(\alpha)^{-1}) + o_{a.s.}(1),$$

where the $o_{a.s.}(1)$ terms are uniform in $1 \le j \le p$. □

Lemma 5.2. *Under the same assumptions in* ***Theorem 2.1****, we have,* $\forall 1 \le j \le p$, $1 \le k \le T-1$, *almost surely,*

$$r_j^t B_j^{-1}(\alpha)P_1^k r_j = \frac{1}{T}tr\left(B^{-1}(\alpha)P_1^k\right) + o_{a.s.}(1) = o_{a.s.}(1),$$

$$r_j^t\tilde{C}_jB_j^{-1}(\alpha)P_1^k r_j = \frac{1}{T}tr\left(\tilde{C}B^{-1}(\alpha)P_1^k\right) + o_{a.s.}(1) = o_{a.s.}(1),$$

where the $o_{a.s.}(1)$ *terms are uniform in* $1 \le j \le p$.

Proof. Notice that, for $1 \le k \le T-1$,

$$P_1 = \begin{pmatrix} \mathbf{0} & 0 \\ \mathbf{I}_{T-1} & \mathbf{0}\end{pmatrix}, \quad P_1^k = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{T-k} & \mathbf{0}\end{pmatrix}, \quad P_1^T = \mathbf{0}, \quad s_j = P_1 r_j.$$

Here P_1^T represents the Tth power of the $T\times T$ matrix P_1, we use P_1^t to denote the transpose of matrix P_1. Denote, for $1\le k\le T$,

$$\frac{1}{T}tr\left(B^{-1}(\alpha)\right):=x_0,\quad \frac{1}{T}tr\left(B^{-1}(\alpha)C\right)=\frac{1}{T}tr\left(\tilde{C}B^{-1}(\alpha)\right):=y_0,$$

$$\frac{1}{T}tr\left(B^{-1}(\alpha)P_1^k\right):=x_k,\quad \frac{1}{T}tr\left(\tilde{C}B^{-1}(\alpha)P_1^k\right):=y_k.$$

It's easy to see that

$$x_T=y_T=0.$$

In addition, since

$$\begin{aligned}&(BA-\alpha\mathbf{I})\left(\frac{1}{\alpha}B\left(AB-\alpha\mathbf{I}\right)^{-1}A-\frac{1}{\alpha}\mathbf{I}\right)\\ =&\frac{1}{\alpha}BAB\left(AB-\alpha\mathbf{I}\right)^{-1}A-B\left(AB-\alpha\mathbf{I}\right)^{-1}A-\frac{1}{\alpha}BA+\mathbf{I}\\ =&\frac{1}{\alpha}B\left(\mathbf{I}+\alpha\left(\mathbf{AB}-\alpha\mathbf{I}\right)^{-\mathbf{1}}\right)A-B\left(AB-\alpha\mathbf{I}\right)^{-1}A-\frac{1}{\alpha}BA+\mathbf{I}=\mathbf{I},\end{aligned}$$

the following equation holds

$$B(AB-\alpha\mathbf{I})^{-\mathbf{1}}\mathbf{A}=\mathbf{I}+\alpha(\mathbf{BA}-\alpha\mathbf{I})^{-\mathbf{1}},\tag{5.5}$$

then we have, for any $1\le j\le p$,

$$\begin{aligned}s_j^t\tilde{C}_jB_j^{-1}(\alpha)C_jr_j&=s_j^t\tilde{C}_j\left(C_j\tilde{C}_j-\alpha\mathbf{I_T}\right)^{-1}C_jr_j\\&=\alpha\cdot s_j^t\left(\tilde{C}_jC_j-\alpha\mathbf{I_T}\right)^{-1}r_j+s_j^tr_j+o_{a.s.}(1)\\&=\alpha\cdot r_j^t\left(C_j\tilde{C}_j-\alpha\mathbf{I_T}\right)^{-1}s_j+o_{a.s.}(1)\\&=\alpha\frac{1}{T}tr(B^{-1}(\alpha)P_1)+o_{a.s.}(1)=\alpha x_1+o_{a.s.}(1).\end{aligned}$$

Now we can derive the recursion equations between x_k and y_k.

Firstly, for x_k, $1\le k\le T-1$, since

$$P_1^k=\left(\sum_{j=1}^p s_js_j^t\sum_{j=1}^p r_jr_j^t\right)B^{-1}(\alpha)P_1^k-\alpha B^{-1}(\alpha)P_1^k,$$

taking the trace and dividing by T on both sides of the equation, we get

$$\begin{aligned}
&\alpha\cdot\frac{1}{T}tr\left(B^{-1}(\alpha)P_1^k\right)\\
=&\frac{1}{T}\sum_{j=1}^p s_j^t\tilde{C}B^{-1}(\alpha)P_1^k s_j=\frac{1}{T}\sum_{j=1}^p s_j^t\tilde{C}\left(B_j(\alpha)+C_jr_jr_j^t+s_js_j^t\tilde{C}\right)^{-1}P_1^k s_j\\
=&\frac{1}{T}\sum_{j=1}^p\frac{s_j^t\tilde{C}\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}P_1^k s_j}{1+s_j^t\tilde{C}\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}s_j}\\
=&\frac{1}{T}\sum_{j=1}^p\frac{s_j^t\tilde{C}_j\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}P_1^k s_j}{1+s_j^t\tilde{C}_j\left(B_j(\alpha)+C_jr_jr_j^t\right)^{-1}s_j}+o_{a.s.}(1)\\
=&\frac{1}{T}\sum_{j=1}^p\frac{1+y_0}{(1+y_0)^2-\alpha x_1^2}\left[s_j^t\tilde{C}_jB_j^{-1}(\alpha)P_1^k s_j-\frac{s_j^t\tilde{C}_jB_j^{-1}(\alpha)C_jr_jr_j^tB_j^{-1}(\alpha)P_1^k s_j}{1+r_j^tB_j^{-1}(\alpha)C_jr_j}\right]+o_{a.s.}(1)\\
=&\frac{p}{T}\frac{1+y_0}{(1+y_0)^2-\alpha x_1^2}\left[\frac{1}{T}tr(\tilde{C}B^{-1}(\alpha)P_1^k)-\frac{\alpha x_1}{1+y_0}\cdot\frac{1}{T}tr\left(B^{-1}(\alpha)P_1^{k+1}\right)\right]+o_{a.s.}(1),
\end{aligned}$$

i.e.

$$\alpha x_k=\frac{p}{T}\cdot\frac{1+y_0}{(1+y_0)^2-\alpha x_1^2}\cdot y_k-\frac{p}{T}\cdot\frac{\alpha x_1}{(1+y_0)^2-\alpha x_1^2}\cdot x_{k+1}+o_{a.s.}(1),\ 1\le k\le T-1. \tag{5.6}$$

In particular, for $k=T-1$, we have

$$\alpha x_{T-1}=\frac{p}{T}\cdot\frac{1+y_0}{(1+y_0)^2-\alpha x_1^2}\cdot y_{T-1}+o_{a.s.}(1). \tag{5.7}$$

Similarly, for y_k, $1\le k\le T$,

$$\begin{aligned}
y_k&=\frac{1}{T}tr\left(\tilde{C}B^{-1}(\alpha)P_1^k\right)\\
&=\frac{1}{T}tr\left(\sum_{j=1}^p r_jr_j^tB^{-1}(\alpha)P_1^k\right)=\frac{1}{T}\sum_{j=1}^p r_j^tB^{-1}(\alpha)P_1^kr_j\\
&=\frac{1}{T}\sum_{j=1}^p\frac{r_j^t\left(B_j(\alpha)+s_js_j^t\tilde{C}_j\right)^{-1}P_1^kr_j}{1+r_j^t\left(B_j(\alpha)+s_js_j^t\tilde{C}_j\right)^{-1}C_jr_j}+o_{a.s.}(1)\\
&=\frac{1}{T}\sum_{j=1}^p\frac{1+y_0}{(1+y_0)^2-\alpha x_1^2}\cdot\left[r_j^tB_j^{-1}(\alpha)P_1^kr_j-\frac{r_j^tB_j^{-1}(\alpha)s_js_j^t\tilde{C}_jB_j^{-1}(\alpha)P_1^kr_j}{1+s_j^t\tilde{C}_jB_j^{-1}(\alpha)s_j}\right]+o_{a.s.}(1)\\
&=\frac{p}{T}\cdot\frac{1+y_0}{(1+y_0)^2-\alpha x_1^2}\cdot\left[\frac{1}{T}tr(B^{-1}(\alpha)P_1^k)-\frac{x_1}{1+y_0}\cdot\frac{1}{T}tr\left(\tilde{C}B^{-1}(\alpha)P_1^{k-1}\right)\right]+o_{a.s.}(1),
\end{aligned}$$

i.e.

$$y_k=\frac{p}{T}\cdot\frac{1+y_0}{(1+y_0)^2-\alpha x_1^2}\cdot x_k-\frac{p}{T}\cdot\frac{x_1}{(1+y_0)^2-\alpha x_1^2}\cdot y_{k-1}+o_{a.s.}(1),\ 1\le k\le T-1. \tag{5.8}$$

Particularly, for $k=T$, we have

$$y_T=\frac{p}{T}\cdot\frac{1+y_0}{(1+y_0)^2-\alpha x_1^2}\cdot x_T-\frac{p}{T}\cdot\frac{x_1}{(1+y_0)^2-\alpha x_1^2}\cdot y_{T-1}+o_{a.s.}(1). \tag{5.9}$$

Note that

$$x_T=y_T=0,$$

so we have either $x_1 = o_{a.s.}(1)$ or $y_{T-1} = o_{a.s.}(1)$.

If $x_1 = o_{a.s.}(1)$, let $(k = 1)$ in equation (5.8), then we have $y_1 = o_{a.s.}(1)$, we denote it by

$$x_1 = o_{a.s.}(1) \xrightarrow[(k=1)]{(5.8)} y_1 = o_{a.s.}(1),$$

consecutively, we have

$$y_1 = o_{a.s.}(1) \xrightarrow[(k=1)]{(5.6)} x_2 = o_{a.s.}(1) \xrightarrow[(k=2)]{(5.8)} y_2 = o_{a.s.}(1) \xrightarrow[(k=2)]{(5.6)} x_3 = o_{a.s.}(1) \xrightarrow[(k=3)]{(5.8)} y_3 = o_{a.s.}(1),$$

Then, recursively, we have for all $1 \le k \le T-1$,

$$x_k = y_k = o_{a.s.}(1).$$

On the other hand, if $y_{T-1} = o_{a.s.}(1)$, since $x_T = y_T = 0$, let $(k = T)$ in equation (5.8), then

$$y_{T-1} = o_{a.s.}(1) \xrightarrow[(k=T-1)]{(5.6)} x_{T-1} = o_{a.s.}(1) \xrightarrow[(k=T-1)]{(5.8)} y_{T-2} = o_{a.s.}(1) \xrightarrow[(k=T-2)]{(5.6)} x_{T-2} = o_{a.s.}(1) \xrightarrow[(k=T-2)]{(5.8)} y_{T-3} = o_{a.s.}(1),$$

Therefore, recursively, we still have for all $1 \le k \le T-1$,

$$x_k = y_k = o_{a.s.}(1).$$

Thus we have, $\forall 1 \le j \le p$, $1 \le k \le T-1$, almost surely,

$$r_j^t B_j^{-1}(\alpha) P_1^k r_j = \frac{1}{T} tr\left(B^{-1}(\alpha) P_1^k\right) + o_{a.s.}(1) = o_{a.s.}(1),$$

$$r_j^t \tilde{C}_j B_j^{-1}(\alpha) P_1^k r_j = \frac{1}{T} tr\left(\tilde{C} B^{-1}(\alpha) P_1^k\right) + o_{a.s.}(1) = o_{a.s.}(1),$$

where the $o_{a.s.}(1)$ terms are uniform in $1 \le j \le p$. □

Lemma 5.3. *Extension of Lemma 5.2 to time lag τ:*

we have, $\forall 1 \le j \le p$, $1 \le k \le [\frac{T}{\tau}]$, almost surely,

$$r_j^t B_j^{-1}(\alpha)(P_1^\tau)^k r_j = \frac{1}{T} tr\left(B^{-1}(\alpha)(P_1^\tau)^k\right) + o_{a.s.}(1) = o_{a.s.}(1),$$

$$r_j^t \tilde{C}_j B_j^{-1}(\alpha)(P_1^\tau)^k r_j = \frac{1}{T} tr\left(\tilde{C} B^{-1}(\alpha)(P_1^\tau)^k\right) + o_{a.s.}(1) = o_{a.s.}(1),$$

where the $o_{a.s.}(1)$ terms are uniform in $1 \le j \le p$.

Proof.

Denote, for $1 \le k \le [\frac{T}{\tau}]$,

$$\frac{1}{T} tr\left(B^{-1}(\alpha)\right) := x_0, \quad \frac{1}{T} tr\left(B^{-1}(\alpha) C\right) = \frac{1}{T} tr\left(\tilde{C} B^{-1}(\alpha)\right) := y_0,$$

$$\frac{1}{T} tr\left(B^{-1}(\alpha)(P_1^\tau)^k\right) := x_k, \quad \frac{1}{T} tr\left(\tilde{C} B^{-1}(\alpha)(P_1^\tau)^k\right) := y_k.$$

It's easy to see that

$$x_{[\frac{T}{\tau}]+1} = y_{[\frac{T}{\tau}]+1} = 0.$$

In addition, for any $1 \le j \le p$,

$$s_j^t \tilde{C}_j B_j^{-1}(\alpha) C_j r_j = \alpha \frac{1}{T} tr(B^{-1}(\alpha) P_1^\tau) + o_{a.s.}(1) = \alpha x_1 + o_{a.s.}(1).$$

Now we can derive the recursion equations between x_k and y_k.

Firstly, for x_k, $1 \le k \le \left[\frac{T}{\tau}\right]$,

$$\alpha \cdot \frac{1}{T} tr\left(B^{-1}(\alpha)(P_1^\tau)^k\right) = o_{a.s.}(1) + \frac{p}{T} \frac{1+y_0}{(1+y_0)^2 - \alpha x_1^2} \left[\frac{1}{T} tr(\tilde{C} B^{-1}(\alpha)(P_1^\tau)^k) - \frac{\alpha x_1}{1+y_0} \cdot \frac{1}{T} tr\left(B^{-1}(\alpha)(P_1^\tau)^{k+1}\right)\right],$$

i.e.

$$\alpha x_k = \frac{p}{T} \cdot \frac{1+y_0}{(1+y_0)^2 - \alpha x_1^2} \cdot y_k - \frac{p}{T} \cdot \frac{\alpha x_1}{(1+y_0)^2 - \alpha x_1^2} \cdot x_{k+1} + o_{a.s.}(1),\ 1 \le k \le \left[\frac{T}{\tau}\right]. \tag{5.10}$$

Similarly, for y_k, $1 \le k \le \left[\frac{T}{\tau}\right] + 1$,

$$\begin{aligned} y_k &= \frac{1}{T} tr\left(\tilde{C} B^{-1}(\alpha)(P_1^\tau)^k\right) \\ &= \frac{p}{T} \cdot \frac{1+y_0}{(1+y_0)^2 - \alpha x_1^2} \cdot \left[\frac{1}{T} tr(B^{-1}(\alpha)(P_1^\tau)^k) - \frac{x_1}{1+y_0} \cdot \frac{1}{T} tr\left(\tilde{C} B^{-1}(\alpha)(P_1^\tau)^{k-1}\right)\right] + o_{a.s.}(1), \end{aligned}$$

i.e.

$$y_k = \frac{p}{T} \cdot \frac{1+y_0}{(1+y_0)^2 - \alpha x_1^2} \cdot x_k - \frac{p}{T} \cdot \frac{x_1}{(1+y_0)^2 - \alpha x_1^2} \cdot y_{k-1} + o_{a.s.}(1),\ 1 \le k \le \left[\frac{T}{\tau}\right] + 1. \tag{5.11}$$

Particularly, for $k = \left[\frac{T}{\tau}\right] + 1$, we have

$$y_{[\frac{T}{\tau}]+1} = \frac{p}{T} \cdot \frac{1+y_0}{(1+y_0)^2 - \alpha x_1^2} \cdot x_{[\frac{T}{\tau}]+1} - \frac{p}{T} \cdot \frac{x_1}{(1+y_0)^2 - \alpha x_1^2} \cdot y_{[\frac{T}{\tau}]} + o_{a.s.}(1). \tag{5.12}$$

Note that

$$x_{[\frac{T}{\tau}]+1} = y_{[\frac{T}{\tau}]+1} = 0,$$

following the same arguments as in Lemma 5.2, we have, $\forall 1 \le j \le p$, $1 \le k \le \left[\frac{T}{\tau}\right]$, almost surely,

$$r_j^t B_j^{-1}(\alpha)(P_1^\tau)^k r_j = \frac{1}{T} tr\left(B^{-1}(\alpha)(P_1^\tau)^k\right) + o_{a.s.}(1) = o_{a.s.}(1),$$

$$r_j^t \tilde{C}_j B_j^{-1}(\alpha)(P_1^\tau)^k r_j = \frac{1}{T} tr\left(\tilde{C} B^{-1}(\alpha)(P_1^\tau)^k\right) + o_{a.s.}(1) = o_{a.s.}(1),$$

where the $o_{a.s.}(1)$ terms are uniform in $1 \le j \le p$.

□

Appendix A: Justification of truncation, centralization and standardization

Recall that $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{pt})^t$, ε_{it} are independent real-valued random variables with $\mathbb{E}(\varepsilon_{it}) = 0, \mathbb{E}\left(|\varepsilon_{it}|^2\right) = 1$, and we are interested in is the LSD of time-lagged covariance matrix

$$A = \frac{1}{T^2}\left(\sum_{i=1}^{T}\varepsilon_i\varepsilon_{i-1}^t\right)\left(\sum_{j=1}^{T}\varepsilon_{j-1}\varepsilon_j^t\right).$$

The assumed moment conditions are: for some constant M, η and arbitrarily small positive δ,

$$\sup_{1\le i\le p, 0\le t\le T}\mathbb{E}\left(|\varepsilon_{it}|^{4+\delta}\right) < M,$$

The aim of the truncation, centralization and standardization procedure is that after these treatment, we may assume that

$$|\varepsilon_{it}| \le \eta T^{1/4}, \quad \mathbb{E}\left(\varepsilon_{it}\right) = 0, \quad Var\left(\varepsilon_{it}\right) = 1, \quad \mathbb{E}\left(|\varepsilon_{it}|^{4+\delta}\right) < M.$$

Since the whole procedure is the same for any time lag τ, we focus on the case of lag-1 sample auto-covariance matrix.

A.1. Truncation

Let $\tilde{\varepsilon}_{jt} = \varepsilon_{jt} I_{(|\varepsilon_{jt}|<\eta T^{1/4})}$, $\tilde{\varepsilon}_t = \left(\tilde{\varepsilon}_{1t}, \cdots, \tilde{\varepsilon}_{pt}\right)^t$, η can be seen as a constant.

Define

$$\tilde{A} = \frac{1}{T^2}\left(\sum_{i=1}^{T}\tilde{\varepsilon}_i\tilde{\varepsilon}_{i-1}^t\right)\left(\sum_{j=1}^{T}\tilde{\varepsilon}_{j-1}\tilde{\varepsilon}_j^t\right),$$

then according to Theorem A.44 of [3] which states that

$$\|F^{AA^*} - F^{BB^*}\| \le \frac{1}{p}\text{rank}\left(A - B\right),$$

we have

$$\begin{aligned}
\|F^A - F^{\tilde{A}}\| &\le \frac{1}{p}\text{rank}\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{\varepsilon}_i\tilde{\varepsilon}_{i-1}^t - \frac{1}{T}\sum_{i=1}^{T}\varepsilon_i\varepsilon_{i-1}^t\right) \\
&\le \frac{1}{p}\text{rank}\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{\varepsilon}_i(\tilde{\varepsilon}_{i-1}^t - \varepsilon_{i-1}^t)\right) + \frac{1}{p}\text{rank}\left(\frac{1}{T}\sum_{i=1}^{T}(\tilde{\varepsilon}_i - \varepsilon_i)\varepsilon_{i-1}^t\right) \\
&\le \frac{1}{p}\sum_{i=1}^{T}\text{rank}\left(\frac{1}{T}\tilde{\varepsilon}_i(\tilde{\varepsilon}_{i-1}^t - \varepsilon_{i-1}^t)\right) + \frac{1}{p}\sum_{i=1}^{T}\text{rank}\left(\frac{1}{T}(\tilde{\varepsilon}_i - \varepsilon_i)\varepsilon_{i-1}^t\right) \\
&\le \frac{2}{p}\sum_{t=0}^{T}\sum_{i=1}^{p} I_{(|\varepsilon_{it}|\ge\eta T^{1/4})},
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{p}\sum_{t=0}^{T}\sum_{i=1}^{p} I_{(|\varepsilon_{it}|\ge\eta T^{1/4})}\right) &\le \frac{1}{p}\sum_{t=0}^{T}\sum_{i=1}^{p}\mathbb{E}\left(\frac{|\varepsilon_{it}|^4}{\eta^4\cdot T} I_{(|\varepsilon_{it}|\ge\eta T^{1/4})}\right) \\
&= \frac{1}{\eta^4 pT}\sum_{i=1}^{p}\sum_{t=0}^{T}\mathbb{E}\left(|\varepsilon_{it}|^4 I_{(|\varepsilon_{it}|\ge\eta T^{1/4})}\right) = o\left(1\right),
\end{aligned}$$

$$Var\left(\frac{1}{p}\sum_{t=0}^{T}\sum_{i=1}^{p}I_{(|\varepsilon_{it}|\geq\eta T^{1/4})}\right)=\frac{1}{p^2}\sum_{t=0}^{T}\sum_{i=1}^{p}Var\left(I_{(|\varepsilon_{it}|\geq\eta T^{1/4})}\right)$$
$$\leq\frac{1}{p^2}\sum_{t=0}^{T}\sum_{i=1}^{p}\mathbb{E}\left(I_{(|\varepsilon_{it}|\geq\eta T^{1/4})}\right)=o\left(\frac{1}{T}\right).$$

Applying Bernstein's inequality

$$\mathbb{P}\left(|S_n|\geq\varepsilon\right)\leq 2\exp\left(-\frac{\varepsilon^2}{2\left(B_n^2+b\varepsilon\right)}\right),$$

where $S_n=\sum_{i=1}^{n}X_i$, $B_n^2=\mathbb{E}S_n^2$, X_i are i.i.d. bounded by b, we can get that, for any small $\varepsilon>0$,

$$\mathbb{P}\left(\frac{1}{p}\sum_{t=0}^{T}\sum_{i=1}^{p}I_{(|\varepsilon_{it}|\geq\eta T^{1/4})}\geq\varepsilon\right)\leq 2\exp\left(-\frac{\varepsilon^2}{2\left(\frac{\varepsilon}{p}+o\left(\frac{1}{T}\right)\right)}\right)=2\exp\left(-K_\varepsilon p\right),$$

which is summable, then by Borel-Cantelli lemma,

$$a.s.\|F^{A}-F^{\tilde{A}}\|\to 0,\text{as } T\to\infty.$$

A.2. Centralization

Let $\hat{\varepsilon}_{it}=\tilde{\varepsilon}_{it}-\mathbb{E}\left(\tilde{\varepsilon}_{it}\right)$, $\hat{\varepsilon}_t=\left(\hat{\varepsilon}_{1t},\cdots,\hat{\varepsilon}_{pt}\right)$, $\hat{A}=\frac{1}{T^2}\left(\sum_{i=1}^{T}\hat{\varepsilon}_i\hat{\varepsilon}_{i-1}^t\right)\left(\sum_{j=1}^{T}\hat{\varepsilon}_{j-1}\hat{\varepsilon}_j^t\right)$.

With Corollary A.42 of [3],

$$L^4\left(F^{AA^*},F^{BB^*}\right)\leq\frac{2}{p^2}tr\left(AA^*+BB^*\right)tr\left(\left(A-B\right)\left(A-B\right)^*\right),$$

we have

$$L^4\left(F^{\hat{A}},F^{\tilde{A}}\right)\leq\frac{2}{p^2}tr\left(\frac{1}{T^2}\left(\sum_{i=1}^{T}\hat{\varepsilon}_i\hat{\varepsilon}_{i-1}^t\right)\left(\sum_{j=1}^{T}\hat{\varepsilon}_{j-1}\hat{\varepsilon}_j^t\right)+\frac{1}{T^2}\left(\sum_{i=1}^{T}\tilde{\varepsilon}_i\tilde{\varepsilon}_{i-1}^t\right)\left(\sum_{j=1}^{T}\tilde{\varepsilon}_{j-1}\tilde{\varepsilon}_j^t\right)\right)$$
$$\cdot\, tr\left(\frac{1}{T^2}\left(\sum_{i=1}^{T}\hat{\varepsilon}_i\hat{\varepsilon}_{i-1}^t-\sum_{i=1}^{T}\tilde{\varepsilon}_i\tilde{\varepsilon}_{i-1}^t\right)\left(\sum_{j=1}^{T}\hat{\varepsilon}_{j-1}\hat{\varepsilon}_j^t-\sum_{j=1}^{T}\tilde{\varepsilon}_{j-1}\tilde{\varepsilon}_j^t\right)\right)$$
$$:=N_1\cdot N_2.$$

For N_2,

$$\begin{aligned}N_2 &= tr\left(\frac{1}{T^2}\left(\sum_{i=1}^{T}\hat{\varepsilon}_i\hat{\varepsilon}_{i-1}^t - \sum_{i=1}^{T}\tilde{\varepsilon}_i\tilde{\varepsilon}_{i-1}^t\right)\left(\sum_{j=1}^{T}\hat{\varepsilon}_{j-1}\hat{\varepsilon}_j^t - \sum_{j=1}^{T}\tilde{\varepsilon}_{j-1}\tilde{\varepsilon}_j^t\right)\right)\\
&= tr\left(\frac{1}{T^2}\sum_{i=1}^{T}\left(\mathbb{E}\left(\tilde{\varepsilon}_i\right)\mathbb{E}\left(\tilde{\varepsilon}_{i-1}^t\right) - \mathbb{E}\left(\tilde{\varepsilon}_i\right)\tilde{\varepsilon}_{i-1}^t - \tilde{\varepsilon}_i\mathbb{E}\left(\tilde{\varepsilon}_{i-1}^t\right)\right)\right.\\
&\quad\left.\cdot\sum_{i=1}^{T}\left(\mathbb{E}\left(\tilde{\varepsilon}_i\right)\mathbb{E}\left(\tilde{\varepsilon}_{i-1}^t\right) - \mathbb{E}\left(\tilde{\varepsilon}_i\right)\tilde{\varepsilon}_{i-1}^t - \tilde{\varepsilon}_i\mathbb{E}\left(\tilde{\varepsilon}_{i-1}^t\right)\right)^t\right)\\
&= \left\|\frac{1}{T}\sum_{i=1}^{T}\left(\mathbb{E}\left(\tilde{\varepsilon}_i\right)\mathbb{E}\left(\tilde{\varepsilon}_{i-1}^t\right) - \mathbb{E}\left(\tilde{\varepsilon}_i\right)\tilde{\varepsilon}_{i-1}^t - \tilde{\varepsilon}_i\mathbb{E}\left(\tilde{\varepsilon}_{i-1}^t\right)\right)\right\|_F^2\\
&\le 2\left\|\frac{1}{T}\sum_{i=1}^{T}\mathbb{E}\left(\tilde{\varepsilon}_i\right)\mathbb{E}\left(\tilde{\varepsilon}_{i-1}^t\right)\right\|_F^2 + 2\left\|\frac{1}{T}\sum_{i=1}^{T}\mathbb{E}\left(\tilde{\varepsilon}_i\right)\tilde{\varepsilon}_{i-1}^t\right\|_F^2 + 2\left\|\frac{1}{T}\sum_{i=1}^{T}\tilde{\varepsilon}_i\mathbb{E}\left(\tilde{\varepsilon}_{i-1}^t\right)\right\|_F^2, \qquad \text{(A.1)}\end{aligned}$$

where $\|\cdot\|_F$ represents for the Frobenius norm of a matrix. Consider the second term, we have

$$\begin{aligned}&\left\|\frac{1}{T}\sum_{i=1}^{T}\mathbb{E}\left(\tilde{\varepsilon}_i\right)\tilde{\varepsilon}_{i-1}^t\right\|_F^2 = \frac{1}{T^2}\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\tilde{\varepsilon}_{j,t-1}\mathbb{E}\left(\tilde{\varepsilon}_{it}\right)\right)^2\\
=&\frac{1}{T^2}\sum_{i,j=1}^{p}\sum_{t=1}^{T}\tilde{\varepsilon}_{j,t-1}^2\left(\mathbb{E}\left(\tilde{\varepsilon}_{it}\right)\right)^2 + \frac{1}{T^2}\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}\tilde{\varepsilon}_{j,t_1-1}\tilde{\varepsilon}_{j,t_2-1}\mathbb{E}\left(\tilde{\varepsilon}_{it_1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{it_2}\right)\\
=:&M_1 + M_2.\end{aligned}$$

Notice that $\sup_{1\le i\le p,1\le t\le T}\mathbb{E}\left(\varepsilon_{it}^{4+\delta}\right) < M$, we have $\frac{1}{\eta^4 pT}\sum_{i=1}^{p}\sum_{t=0}^{T}\mathbb{E}\left(|\varepsilon_{it}|^4 I_{(|\varepsilon_{it}|\ge\eta T^{1/4})}\right) = o(1)$, then

$$\begin{aligned}\mathbb{E}\left(M_1\right) &= \frac{1}{T^2}\sum_{i,j=1}^{p}\sum_{t=1}^{T}\mathbb{E}\left(\tilde{\varepsilon}_{j,t-1}^2\right)\left(\mathbb{E}\left(\tilde{\varepsilon}_{it}\right)\right)^2\\
&\le \frac{C_1}{T^2}\sum_{i,j=1}^{p}\sum_{t=1}^{T}\left(\mathbb{E}\left(|\varepsilon_{it}|I_{(|\varepsilon_{it}|\ge\eta T^{1/4})}\right)\right)^2\\
&\le \frac{C_1}{T^2}\sum_{i,j=1}^{p}\sum_{t=1}^{T}\frac{1}{\eta^6\cdot T^{3/2}}\left(\mathbb{E}\left(|\varepsilon_{it}|^4 I_{(|\varepsilon_{it}|\ge\eta T^{1/4})}\right)\right)^2\\
&= O\left(T^{-\frac{1}{2}}\right),\end{aligned}$$

Moreover,

$$\begin{aligned}Var\left(M_1\right) &= \frac{1}{T^4}\sum_{j=1}^{p}\sum_{t=1}^{T}\mathbb{E}\left(\tilde{\varepsilon}_{j,t-1}^2 - \mathbb{E}\left(\tilde{\varepsilon}_{j,t-1}^2\right)\right)^2\left(\sum_{i=1}^{p}\left(\mathbb{E}\left(\tilde{\varepsilon}_{it}\right)\right)^2\right)^2\\
&\le \frac{1}{T^4}\sum_{j=1}^{p}\sum_{t=1}^{T}\mathbb{E}\left(\tilde{\varepsilon}_{j,t-1}^2\right)^4\left(\sum_{i=1}^{p}\left(\mathbb{E}\left(|\varepsilon_{it}|I_{(|\varepsilon_{it}|\ge\eta\cdot T^{1/4})}\right)\right)^2\right)^2\\
&\le \frac{C_2}{T^4}\sum_{j=1}^{p}\sum_{t=1}^{T}\frac{1}{T^3}\left(\sum_{i=1}^{p}\left(\mathbb{E}\left(|\varepsilon_{it}|^4 I_{(|\varepsilon_{it}|\ge\eta\cdot T^{1/4})}\right)\right)^2\right)^2 = O\left(T^{-3}\right).\end{aligned}$$

Therefore, $a.s. \quad M_1 \to 0, \text{as } T \to \infty$.

For the term M_2, we have

$$\begin{aligned}
\mathbb{E}(M_2) &= \frac{1}{T^2}\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}\mathbb{E}\left(\tilde{\varepsilon}_{j,t_1-1}\tilde{\varepsilon}_{j,t_2-1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{it_1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{it_2}\right)\\
&= \frac{1}{T^2}\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}\mathbb{E}\left(\tilde{\varepsilon}_{j,t_1-1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{j,t_2-1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{it_1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{it_2}\right)\\
&\le \frac{1}{T^2}\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}\frac{1}{\eta^{12}\cdot T^3}\left(\sup_{1\le i\le p,0\le t\le T}\mathbb{E}\left(|\varepsilon_{it}|^4 I_{(|\varepsilon_{it}|\ge\eta\cdot T^{1/4})}\right)\right)^4 = O\left(T^{-1}\right),
\end{aligned}$$

$$\begin{aligned}
Var(M_2) &= \frac{1}{T^4}\sum_{j=1}^{p}\sum_{t_1\neq t_2}Var\left(\tilde{\varepsilon}_{j,t_1-1}\tilde{\varepsilon}_{j,t_2-1}\right)\left(\sum_{i=1}^{p}\mathbb{E}\left(\tilde{\varepsilon}_{it_1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{it_2}\right)\right)^2\\
&\le \frac{1}{T^4}\sum_{j=1}^{p}\sum_{t_1\neq t_2}\mathbb{E}\left(\tilde{\varepsilon}^2_{j,t_1-1}\right)\mathbb{E}\left(\tilde{\varepsilon}^2_{j,t_2-1}\right)\left(\sum_{i=1}^{p}\left(\sup_{1\le i\le p,0\le t\le T}\mathbb{E}\left(\tilde{\varepsilon}_{it}\right)\right)^2\right)^2\\
&\le \frac{C_3}{T^4}\sum_{j=1}^{p}\sum_{t_1\neq t_2}\frac{1}{T^3}\left(\sum_{i=1}^{p}\left(\sup_{1\le i\le p,0\le t\le T}\mathbb{E}\left(|\varepsilon_{it}|^4 I_{(|\varepsilon_{it}|\ge\eta\cdot T^{1/4})}\right)\right)^2\right)^2 = O\left(T^{-2}\right).
\end{aligned}$$

Therefore, a.s. $M_2 \to 0$, as $T \to \infty$.

Consequently, $\left\|\frac{1}{T}\sum_{i=1}^{T}\mathbb{E}\left(\tilde{\varepsilon}_i\right)\tilde{\varepsilon}^t_{i-1}\right\|_F^2 \to 0, a.s.$ Similarly, we can prove that the last term in equation (A.1) tends to zero almost surely. As for the first term, we have

$$\begin{aligned}
\left\|\frac{1}{T}\sum_{i=1}^{T}\mathbb{E}\left(\tilde{\varepsilon}_i\right)\mathbb{E}\left(\tilde{\varepsilon}^t_{i-1}\right)\right\|_F^2 &= \sum_{i,j=1}^{p}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left(\tilde{\varepsilon}_{it}\right)\mathbb{E}\left(\tilde{\varepsilon}_{j,t-1}\right)\right)\right)^2\\
&= \frac{1}{T^2}\sum_{i,j=1}^{p}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathbb{E}\left(\tilde{\varepsilon}_{it_1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{j,t_1-1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{it_2}\right)\mathbb{E}\left(\tilde{\varepsilon}_{j,t_2-1}\right)\\
&\le \frac{C_4}{T^2}\sum_{i,j=1}^{p}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\frac{1}{T^3}\left(\sup_{1\le i\le p,0\le t\le T}\mathbb{E}\left(|\varepsilon_{it}|^4 I_{(|\varepsilon_{it}|\ge\eta\cdot T^{1/4})}\right)\right)^4 = O\left(T^{-1}\right).
\end{aligned}$$

Therefore

$$N_2 = tr\left(\frac{1}{T^2}\left(\sum_{i=1}^{T}\hat{\varepsilon}_i\hat{\varepsilon}^t_{i-1} - \sum_{i=1}^{T}\tilde{\varepsilon}_i\tilde{\varepsilon}^t_{i-1}\right)\left(\sum_{j=1}^{T}\hat{\varepsilon}_{j-1}\hat{\varepsilon}^t_j - \sum_{j=1}^{T}\tilde{\varepsilon}_{j-1}\tilde{\varepsilon}^t_j\right)\right) \to 0, a.s.$$

Now, we consider N_1,

$$\frac{1}{p^2}tr\left(\frac{1}{T^2}\left(\sum_{i=1}^{T}\hat{\varepsilon}_i\hat{\varepsilon}^t_{i-1}\right)\left(\sum_{j=1}^{T}\hat{\varepsilon}_{j-1}\hat{\varepsilon}^t_j\right) + \frac{1}{T^2}\left(\sum_{i=1}^{T}\tilde{\varepsilon}_i\tilde{\varepsilon}^t_{i-1}\right)\left(\sum_{j=1}^{T}\tilde{\varepsilon}_{j-1}\tilde{\varepsilon}^t_j\right)\right) =: M_3 + M_4,$$

Firstly, for M_3, since $\mathbb{E}\left(\hat{\varepsilon}_{it}\right) = 0$,

$$\begin{aligned}
\mathbb{E}(M_3) &= \mathbb{E}\left(\frac{1}{p^2T^2}\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\hat{\varepsilon}_{it}\hat{\varepsilon}_{j,t-1}\right)^2\right)\\
&= \frac{1}{p^2T^2}\sum_{i,j=1}^{p}\sum_{t=1}^{T}\mathbb{E}\left(\hat{\varepsilon}^2_{it}\right)\mathbb{E}\left(\hat{\varepsilon}^2_{j,t-1}\right) = O\left(\frac{1}{T}\right).
\end{aligned}$$

Moreover,

$$\begin{aligned}
Var\left(M_3\right) &= \mathbb{E}\left(\frac{1}{p^2T^2}\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\hat{\varepsilon}_{it}\hat{\varepsilon}_{j,t-1}\right)^2\right)^2 - \left(\mathbb{E}\left(M_3\right)\right)^2 \\
&= \frac{1}{p^4T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{j,t-1}^2\right)^2 + \frac{1}{p^4T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}\hat{\varepsilon}_{it_1}\hat{\varepsilon}_{j,t_1-1}\hat{\varepsilon}_{it_2}\hat{\varepsilon}_{j,t_2-1}\right)^2 + O\left(\frac{1}{T^2}\right) \\
&\leq O\left(\frac{1}{T^2}\right) + O\left(\frac{1}{T^3}\right) + O\left(\frac{1}{T^2}\right) = O\left(\frac{1}{T^2}\right).
\end{aligned}$$

Therefore $M_3 \to 0$, a.s. Next for M_4,

$$\begin{aligned}
\mathbb{E}\left(M_4\right) &= \mathbb{E}\left(\frac{1}{p^2T^2}\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{j,t-1}\right)^2\right) \\
&= \frac{1}{p^2T^2}\sum_{i,j=1}^{p}\sum_{t=1}^{T}\mathbb{E}\tilde{\varepsilon}_{it}^2\mathbb{E}\tilde{\varepsilon}_{j,t-1}^2 + \frac{1}{p^2T^2}\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}\mathbb{E}\left(\tilde{\varepsilon}_{it_1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{j,t_1-1}\right)\mathbb{E}\left(\tilde{\varepsilon}_{it_2}\right)\mathbb{E}\left(\tilde{\varepsilon}_{j,t_2-1}\right) \\
&\leq O\left(\frac{1}{T}\right) + \frac{1}{p^2T^2}\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}\frac{1}{\eta^{12}T^3}\left(\sup_{1\leq i\leq p, 0\leq t\leq T}\mathbb{E}\left(|\varepsilon_{it}|^4 I_{(|\varepsilon_{it}|\geq\eta\cdot T^{1/4})}\right)\right)^4 = O\left(\frac{1}{T}\right).
\end{aligned}$$

$$\begin{aligned}
Var\left(M_4\right) &= \frac{1}{p^4T^4}Var\left(\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{j,t-1}\right)^2\right) \\
&\leq \frac{1}{p^4T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{j,t-1}\right)^2\right)^2 \\
&= \frac{1}{p^4T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t=1}^{T}\tilde{\varepsilon}_{it}^2\tilde{\varepsilon}_{j,t-1}^2\right)^2 + \frac{1}{p^4T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}\tilde{\varepsilon}_{it_1}\tilde{\varepsilon}_{j,t_1-1}\tilde{\varepsilon}_{it_2}\tilde{\varepsilon}_{j,t_2-1}\right)^2 \\
&\leq O\left(\frac{1}{T^2}\right) + O\left(\frac{1}{T^6}\right) = O\left(\frac{1}{T^2}\right).
\end{aligned}$$

Therefore, $M_4 \to 0, a.s.$. All in all,

$$L^4\left(F^{\hat{A}}, F^{\tilde{A}}\right) \leq N_1 \cdot N_2 \leq 4\left(M_3 + M_4\right)\left(M_1 + M_2\right) \to 0, a.s. T \to \infty.$$

A.3. Rescaling

Define $\tilde{\varepsilon}_{it} = \varepsilon_{it} I_{\{|\varepsilon_{it}|\leq\eta T^{1/4}\}}$, $\hat{\varepsilon}_{it} = \tilde{\varepsilon}_{it} - \mathbb{E}\tilde{\varepsilon}_{it}$, $\hat{\sigma}_{it}^2 = \mathbb{E}|\hat{\varepsilon}_{it}|^2 = \mathbb{E}|\tilde{\varepsilon}_{it} - \mathbb{E}\tilde{\varepsilon}_{it}|^2$ and $\overline{\varepsilon}_{it} = \frac{\hat{\varepsilon}_{it}}{\hat{\sigma}_{it}}$, we first show that $\hat{\sigma}_{it}^2 s$ tend to 1 uniformly.

We consider the distance between $\hat{A} = \frac{1}{T^2}\left(\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}^t\right)\left(\sum_{t=1}^{T}\hat{\varepsilon}_{t-1}\hat{\varepsilon}_t^t\right)$ and $\overline{A} = \frac{1}{T^2}\left(\sum_{t=1}^{T}\overline{\varepsilon}_t\overline{\varepsilon}_{t-1}^t\right)\left(\sum_{t=1}^{T}\overline{\varepsilon}_{t-1}\overline{\varepsilon}_t^t\right)$.

Since $\varepsilon_{it} = \tilde{\varepsilon}_{it} + \varepsilon_{it} I_{\{|\varepsilon_{it}| > \eta T^{1/4}\}} := \tilde{\varepsilon}_{it} + r_{it}$, we have $0 = \mathbb{E}(\tilde{\varepsilon}_{it} + r_{it}) = \mathbb{E}(\tilde{\varepsilon}_{it}) + \mathbb{E}(r_{it})$. Next,

$$\begin{aligned} 1 = Var(\varepsilon_{it}) = Var(\tilde{\varepsilon}_{it} + r_{it}) &= Var(\tilde{\varepsilon}_{it}) + Var(r_{it}) + 2Cov(\tilde{\varepsilon}_{it}, r_{it}) \\ &= Var(\tilde{\varepsilon}_{it}) + Var(r_{it}) + 2\left[\mathbb{E}(\tilde{\varepsilon}_{it} r_{it}) - \mathbb{E}(\tilde{\varepsilon}_{it})\mathbb{E}(r_{it})\right] \\ &= Var(\tilde{\varepsilon}_{it}) + Var(r_{it}) + 2\left[\mathbb{E}(r_{it})\right]^2, \end{aligned}$$

so that

$$1 - Var(\tilde{\varepsilon}_{it}) = \mathbb{E}\left(r_{it}^2\right) + \left[\mathbb{E}(r_{it})\right]^2 \le 2\mathbb{E}\left(r_{it}^2\right).$$

It follows that

$$\begin{aligned} \max_{1\le i\le p, 0\le t\le T} \left(1 - \hat{\sigma}_{it}^2\right) &\le 2 \max_{1\le i\le p, 0\le t\le T} \mathbb{E}\left(\varepsilon_{it}^2 I_{\{|\varepsilon_{it}| > \eta T^{1/4}\}}\right) \\ &\le 2 \max_{1\le i\le p, 0\le t\le T} \mathbb{E}\left(\frac{\varepsilon_{it}^{4+\delta}}{\varepsilon_{it}^{2+\delta}} I_{\{|\varepsilon_{it}| > \eta T^{1/4}\}}\right) \\ &\le 2\frac{1}{\eta^{2+\delta} T^{1/2+\delta/4}} \max_{1\le i\le p, 0\le t\le T} \mathbb{E}\left(\varepsilon_{it}^{4+\delta} I_{\{|\varepsilon_{it}| > \eta T^{1/4}\}}\right) \\ &\le \frac{2M}{\eta^{2+\delta} T^{1/2+\delta/4}} \to 0, \text{ as } T \to \infty, \end{aligned}$$

where the last step uses the uniform bound $\sup_{1\le i\le p, 0\le t\le T} \mathbb{E}\left(|\varepsilon_{it}|^{4+\delta}\right) < M$. As

$$1 - \hat{\sigma}_{it} = \frac{1 - \hat{\sigma}_{it}^2}{1 + \hat{\sigma}_{it}} \le 1 - \hat{\sigma}_{it}^2,$$

we have

$$k_T := \max_{i,t} (1 - \hat{\sigma}_{it}) \to 0, \ 1 - k_T \le \hat{\sigma}_{it} \le 1, \ \forall i, t$$

and

$$0 \le \frac{1}{\hat{\sigma}_{it}} - 1 \le \frac{1}{1 - k_T} - 1 \to 0, \text{ as } T \to \infty.$$

According to Corollary A.42 of [3], we have

$$\begin{aligned} L^4\left(F^{\hat{A}}, F^{\overline{A}}\right) \le \frac{2}{p^2} &\left(\left\|\frac{1}{T}\sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}^t\right\|_F^2 + \left\|\frac{1}{T}\sum_{t=1}^T \overline{\varepsilon}_t \overline{\varepsilon}_{t-1}^t\right\|_F^2\right) \\ &\cdot \left\|\frac{1}{T}\sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}^t - \frac{1}{T}\sum_{t=1}^T \overline{\varepsilon}_t \overline{\varepsilon}_{t-1}^t\right\|_F^2. \end{aligned}$$

Firstly, consider

$$\begin{aligned} \frac{1}{p}\left\|\frac{1}{T}\sum_{t=1}^T \left(\hat{\varepsilon}_t \hat{\varepsilon}_{t-1}^t - \overline{\varepsilon}_t \overline{\varepsilon}_{t-1}^t\right)\right\|_F^2 &= \frac{1}{p}\left\|\frac{1}{T}\sum_{t=1}^T \left(\hat{\varepsilon}_t \hat{\varepsilon}_{t-1}^t - \overline{\varepsilon}_t \hat{\varepsilon}_{t-1}^t + \overline{\varepsilon}_t \hat{\varepsilon}_{t-1}^t - \overline{\varepsilon}_t \overline{\varepsilon}_{t-1}^t\right)\right\|_F^2 \\ &\le 2\left(\frac{1}{p}\left\|\frac{1}{T}\sum_{t=1}^T \left(\hat{\varepsilon}_t \hat{\varepsilon}_{t-1}^t - \overline{\varepsilon}_t \hat{\varepsilon}_{t-1}^t\right)\right\|_F^2 + \frac{1}{p}\left\|\frac{1}{T}\sum_{t=1}^T \left(\overline{\varepsilon}_t \hat{\varepsilon}_{t-1}^t - \overline{\varepsilon}_t \overline{\varepsilon}_{t-1}^t\right)\right\|_F^2\right) \\ &:= 2\left(M_5 + M_6\right), \end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left(M_5^2\right) &= \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\left(1-\frac{1}{\hat{\sigma}_{it}}\right)\hat{\varepsilon}_{it}\hat{\varepsilon}_{j,t-1}\right)^2\right)^2 \\
&= \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t=1}^{T}\left(1-\frac{1}{\hat{\sigma}_{it}}\right)^2\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{j,t-1}^2\right)^2 \\
&+ \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}^{T}\left(1-\frac{1}{\hat{\sigma}_{it_1}}\right)\left(1-\frac{1}{\hat{\sigma}_{it_2}}\right)\hat{\varepsilon}_{it_1}\hat{\varepsilon}_{j,t_1-1}\hat{\varepsilon}_{it_2}\hat{\varepsilon}_{j,t_2-1}\right)^2 \\
&\leq \frac{p^2}{T^2}\max_{i,t}\left(\hat{\sigma}_{it}-1\right)^4+\frac{1}{T^2}\max_{i,t}\left(\hat{\sigma}_{it}-1\right)^4 = O\left(\frac{1}{T^{2+\delta}}\right),
\end{aligned}
$$

Therefore $M_5 \to 0$, a.s.

Similarly for M_6,

$$
\begin{aligned}
\mathbb{E}\left(M_6^2\right) &= \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{it}}\left(1-\frac{1}{\hat{\sigma}_{j,t-1}}\right)\hat{\varepsilon}_{it}\hat{\varepsilon}_{j,t-1}\right)^2\right)^2 \\
&= \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{it}^2}\left(1-\frac{1}{\hat{\sigma}_{j,t-1}}\right)^2\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{j,t-1}^2\right)^2 \\
&+ \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}^{T}\frac{1}{\hat{\sigma}_{it_1}\hat{\sigma}_{it_2}}\left(1-\frac{1}{\hat{\sigma}_{j,t_1-1}}\right)\left(1-\frac{1}{\hat{\sigma}_{j,t_2-1}}\right)\hat{\varepsilon}_{it_1}\hat{\varepsilon}_{j,t_1-1}\hat{\varepsilon}_{it_2}\hat{\varepsilon}_{j,t_2-1}\right)^2 \\
&\leq \frac{p^2}{T^2}\max_{i,t}\left(\hat{\sigma}_{it}-1\right)^4+\frac{1}{T^2}\max_{i,t}\left(\hat{\sigma}_{it}-1\right)^4 = O\left(\frac{1}{T^{2+\delta}}\right),
\end{aligned}
$$

Therefore, $\frac{1}{p}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}^t-\overline{\varepsilon}_t\overline{\varepsilon}_{t-1}^t\right)\right\|_F^2 \to 0$, a.s.

Secondly, consider

$$
\begin{aligned}
&\frac{1}{p}\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}^t\right\|_F^2+\left\|\frac{1}{T}\sum_{t=1}^{T}\overline{\varepsilon}_t\overline{\varepsilon}_{t-1}^t\right\|_F^2\right) \\
&= \frac{1}{p}tr\left(\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}^t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{t-1}\hat{\varepsilon}_t^t\right)\right)+\frac{1}{p}tr\left(\left(\frac{1}{T}\sum_{t=1}^{T}\overline{\varepsilon}_t\overline{\varepsilon}_{t-1}^t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\overline{\varepsilon}_{t-1}\overline{\varepsilon}_t^t\right)\right) \\
&:= M_7+M_8
\end{aligned}
$$

Consider $M_7 = \frac{1}{pT^2}tr\left(\left(\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}^t\right)\left(\sum_{t=1}^{T}\hat{\varepsilon}_{t-1}\hat{\varepsilon}_t^t\right)\right)$,

$$\begin{aligned}\mathbb{E}\left(M_7^2\right) &= \mathbb{E}\left(\frac{1}{pT^2}\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\hat{\varepsilon}_{it}\hat{\varepsilon}_{j,t-1}\right)^2\right)^2 \\ &= \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{j,t-1}^2\right)^2 + \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}\hat{\varepsilon}_{it_1}\hat{\varepsilon}_{j,t_1-1}\hat{\varepsilon}_{it_2}\hat{\varepsilon}_{j,t_2-1}\right)^2 \\ &= O\left(1\right) + O\left(\frac{1}{T^2}\right) = O\left(1\right).\end{aligned}$$

Moreover,

$$\begin{aligned}\mathbb{E}\left(M_8^2\right) &= \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\left(\sum_{t=1}^{T}\frac{\hat{\varepsilon}_{it}}{\hat{\sigma}_{it}}\cdot\frac{\hat{\varepsilon}_{j,t-1}}{\hat{\sigma}_{j,t-1}}\right)^2\right)^2 \\ &= \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t=1}^{T}\frac{\hat{\varepsilon}_{it}^2}{\hat{\sigma}_{it}^2}\cdot\frac{\hat{\varepsilon}_{j,t-1}^2}{\hat{\sigma}_{j,t-1}^2}\right)^2 \\ &\quad + \frac{1}{p^2T^4}\mathbb{E}\left(\sum_{i,j=1}^{p}\sum_{t_1\neq t_2}^{T}\frac{\hat{\varepsilon}_{it_1}}{\hat{\sigma}_{it_1}}\cdot\frac{\hat{\varepsilon}_{j,t_1-1}}{\hat{\sigma}_{j,t_1-1}}\cdot\frac{\hat{\varepsilon}_{it_2}}{\hat{\sigma}_{it_2}}\cdot\frac{\hat{\varepsilon}_{j,t_2-1}}{\hat{\sigma}_{j,t_2-1}}\right) = O\left(1\right).\end{aligned}$$

Therefore

$$\begin{aligned}L^4\left(F^{\hat{A}}, F^{\overline{A}}\right) &\leq 2(M_7+M_8)\cdot 2(M_5+M_6) \\ &= 4(M_7M_5+M_7M_6+M_8M_5+M_8M_6),\end{aligned}$$

since $\mathbb{E}|M_7M_5| \leq \left(\mathbb{E}(M_7^2)\right)^{1/2}\left(\mathbb{E}(M_5^2)\right)^{1/2} = O(\frac{1}{T^{1+\delta/2}})$, we have $M_7M_5 \to 0$, a.s. and similarly for M_7M_6, M_8M_5, M_8M_6, therefore, $L^4\left(F^{\hat{A}}, F^{\overline{A}}\right) \to 0$, a.s.

References

[1] Anderson, G. W., Guionnet, A., and Zeitouni, O. (2010). *An introduction to random matrices*(No. 118). Cambridge University Press.

[2] Bai, Z.D. and Silverstein, J.W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *The Annals of Probability*, Vol. 26, NO. 1, 316-345.

[3] Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices* (2nd ed.) Springer.

[4] Bai, Z.D. and Chen Wang (2014). A note on the limiting spectral distribution of a symmetrized auto-cross covariance matrix, *ArXiv*:1403.2578.

[5] Bai, Z.D. and Yin, Y.Q.(1993). Limit of the smallest eigenvalues of large dimensional covariance matrix. *Ann. Probab.* **21**(3), 1275-1294.

[6] B. Jin, C. Wang, Z. Bai, K. Nair and M. Harding(2013). Limiting Spectral distribution of a symmetrized auto-cross covariance matrix *Ann. Applied Probab.*, **24**(3),1199-1225.

[7] Lam, C., Yao, Q. and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* **98(4)**, 901-918.

[8] Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Ann. Statist.* **40(2)**, 694-726.

[9] Zeng Li, Qinwen Wang and Jianfeg Yao(2014). Identifying the number of factors from singular values of a large sample auto-covariance matrix, *ArXiv*:1410.3687.

[10] Liu, H., Aue, A., and Paul, D. (2013). On the Marcenko-Pastur law for linear time series. *arXiv preprint* arXiv:1310.7270.

[11] Marčenko, V.A. and Pastur, L.A.(1967), Distribution of eigenvalues in certain sets of random matrices. *Math. USSR-Sb.* **1**, 457-483.

[12] Pastur, L. A. and Shcherbina, M. (2010). *Eigenvalue distribution of large random matrices.* American Mathematical Society, Providence, Rhode Island

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

[14] C. Wang, B. Jin, Z. Bai, K. Nair and M. Harding(2014). Strong limit of the extreme eigenvalues of a symmetrized auto-cross covariance matrix. *ArXiv:* 1312.2277.