Limiting Spectral Distribution of Sample Autocovariance Matrices

Anirban Basak *
Department of Statistics
Stanford University
USA
anirbanb@stanford.edu

Arup Bose †
Statistics and Mathematics Unit
Indian Statistical Institute
INDIA
bosearu@gmail.com

Sanchayan Sen ‡
Courant Institute of
Mathematical Sciences
USA
sen@cims.nyu.edu

August 14, 2011

Abstract

The empirical spectral distribution (ESD) of the sample variance covariance matrix of i.i.d. observations under suitable moment conditions converges almost surely as the dimension tends to infinity. The limiting spectral distribution (LSD) is universal and is known in closed form with support $[0,\ 4]$. In this article we show that the ESD of the sample autocovariance matrix converges as the dimension increases, when the time series is a linear process with reasonable restriction on the coefficients. This limit does not depend on the distribution of the underlying driving i.i.d. sequence but in contrast to the sample variance covariance matrix, its support is unbounded. The limit moments are certain functions of the autocovariances. This limit is inconsistent in the sense that it does not coincide with the spectral distribution of the theoretical autocovariance matrix. However, if we consider a suitably tapered version of the autocovariance matrix, then its LSD also exists and is consistent. We also discuss the existence of the LSD for banded sample autocovariance matrices. For banded matrices, the limit has unbounded support as long as the number of nonzero diagonals in proportion to the dimension of the matrix is bounded away from zero. If this ratio tends to zero, then the limit has bounded support. Finally we also study the LSD of a naturally modified version of the autocovariance matrix which is not nonnegative definite.

AMS 2010 Subject Classification Primary 60B20, Secondary 60B10, 60F99, 60G57, 60G10, 62M10.

Key words and phrases. Autocovariance function, autocovariance matrix, linear process, spectral distribution, stationary process, Toeplitz matrix, banded and tapered autocovariance matrix.

*Supported by Melvin and Joan Lane endowed Stanford Graduate Fellowship Fund.

†Research supported by J.C. Bose National Fellowship, Dept. of Science and Technology, Govt. of India. Part of the work was done while this author was visiting Dept. of Economics, Univ. of Cincinnati in 2010.

‡Supported by NYU graduate fellowship under Henry M. MacCracken Program.

1 Introduction

Let $X = \{X_t\}$ be a *stationary* process with $\mathbb{E}(X_t) = 0$ and $\mathbb{E}(X_t^2) < \infty$. The *autocovariance function* $\gamma_X(\cdot)$ and the *autocovariance matrix* $\Sigma_n(X)$ of order n are defined as:

$$\gamma_X(k) = cov(X_0, X_k),\ k = 0, 1, \ldots \quad \text{and} \quad \Sigma_n(X) = ((\gamma_X(i-j)))_{1\le i,j\le n}.$$

These quantities appear frequently in time series analysis. From the spectral representation of autocovariances, there is a distribution F_X, called the *spectral distribution* of $\{\gamma_X(h)\}$ that satisfies

$$\gamma_X(h) = \int_{(0,\ 1]} \exp(2\pi i h x) dF_X(x) \quad \text{for all} \quad h. \tag{1.1}$$

This correspondence between $\gamma(\cdot)$ and F is one to one. It is also known that if $\sum_{k=1}^{\infty} |\gamma_X(k)| < \infty$ then the density (also known as the spectral density of X or $\gamma(\cdot)$) of F_X is given by

$$f_X(t) = \sum_{k=-\infty}^{\infty} \exp(-2\pi i t k)\gamma_X(k),\ t \in (0,\ 1]. \tag{1.2}$$

Now suppose that $A_{n\times n}$ is any real symmetric matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be its eigenvalues. The *Empirical Spectral Distribution (ESD)* of A_n is defined as,

$$F^{A_n}(x) = n^{-1} \sum_{i=1}^{n} \mathbb{I}(\lambda_i \le\ x). \tag{1.3}$$

The *Limiting Spectral Distribution (or measure) (LSD)* F is defined as the weak limit of the sequence $\{F^{A_n}\}$, if it exists. We write $F^{A_n} \xrightarrow{w} F$. The entries of A_n may be random. In that case, the limit is taken to be either in almost sure or in probability sense. There is a growing literature on the study of spectral distribution of random matrices. See Bose, Hazra and Saha (2010) for a recent but restricted review of this area.

Any matrix T_n of the form $((t_{i-j}))_{1\le i,j\le n}$ is a *Toeplitz* matrix and hence $\Sigma_n(X)$ is a Toeplitz matrix. For simplicity suppose that T_n is symmetric (that is $t_k = t_{-k}$ for all k) and $\sum_{k=-\infty}^{\infty} |t_k| < \infty$. From Szego's theory of Toeplitz operators (see for example Böttcher and Silberman (1998)), the LSD of T_n exists and may be described as follows. Define

$$f(x) = \sum_{k=-\infty}^{\infty} t_k \exp(-2\pi i x k),\ x \in (0,\ 1].$$

Then the LSD is the distribution of $f(U)$ where U is uniformly distributed on $(0,\ 1]$. In particular if $\sum_{k=1}^{\infty} |\gamma_X(k)| < \infty$, then the LSD of $\Sigma_n(X)$ exists and equals the distribution of $f_X(U)$. It is interesting to note that

$$\mathbb{E}[f_X(U)]^h \;=\; \mathbb{E}\Big[\gamma_X(0) + \sum_{k=1}^{\infty} \gamma_X(k)\{e^{2\pi i U k} + e^{-2\pi i U k)}\}\Big]^h = \sum_{S_{h,\infty}} \eta_{\mathbf{k}} \prod_{j=0}^{\infty} \gamma_X(j)^{k_j} \tag{1.4}$$

where

$$S_{h,\infty} = \{(k_0 \dots k_d, \dots) : k_j \geq 0 \text{ for all } j,\ k_0 + \dots + k_d + \dots = h\}. \tag{1.5}$$

and $\eta_{\mathbf{k}}$ is number possible ways of choosing $\mathbf{b} \in \{-1, 1\}^h$ such that $\sum_i b_i n_i = 0$ and $b_i = 1$ whenever $n_i = 0$, where $k_j = \#\{i : n_i = j\}$, $n_i \in \mathbb{N}$.
Now let us turn to the sample autocovariance matrix. This is the usual *nonnegative definite* estimate of $\Sigma_n(X)$ and equals

$$\Gamma_n(X) = ((\hat{\gamma}_X(i-j)))_{1 \leq i,j \leq n} \text{ where } \hat{\gamma}_X(k) = n^{-1} \sum_{i=1}^{n-|k|} X_i X_{i+|k|} \tag{1.6}$$

It is also a Toeplitz matrix. Sen (2006) raised the natural question whether the LSD of $\Gamma_n(X)$ exists and if so, whether there is any type of universality of the limit with respect to the distribution of the process that drives X. His simulations suggested that there is convergence as well as universality. Basak (2009) made an initial study in the special case where X is an i.i.d. process. Sen (2010) has some results for the case where X is an $MA(1)$ process. Our primary goal is to study the LSD of this matrix and some of its variants (described later) in the general setup where

$$X_t = \sum_{k=0}^{\infty} \theta_k \varepsilon_{t-k} \tag{1.7}$$

is a linear process and $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ is a sequence of independent random variables with appropriate conditions.

When $\{\varepsilon_t\}$ is i.i.d., for every fixed k, $\hat{\gamma}_X(k) \to \gamma_X(k)$ and this is the basis for the extensive use of the *sample autocovariance* sequence $\{\hat{\gamma}_X(k)\}$ in time series analysis. However, this convergence does not yield the LSD of $\Gamma_n(X)$ to be $f_X(U)$. This is because there is no uniformity in the above convergence. Indeed, $\Sigma_n(X) - \Gamma_n(X)$ does not converge to zero in any reasonable norm. In particular, it is known that the largest eigenvalue or the operator norm of this matrix does not converge to zero. See for example Wu and Pourahmadi (2009), McMurray and Politis (2010) and Xiao and Wu (2011). This *suggests* that even when the LSD of $\Gamma_n(X)$ exists, it may not be the same as the LSD $f_X(U)$ of $\Sigma_n(X)$.

Incidentally, $\Gamma_n(X)$ reminds us of the sample variance covariance matrix, S, whose spectral properties are well studied. We refer to Bai (1999) for some of the basic references. In particular the LSD of S (with i.i.d. entries) under suitable conditions is given by the Marčenko Pastur (1967) law which is supported on the interval $[0, 4]$. Obtaining any result on the limiting spectrum of $\Gamma_n(X)$ does not appear to be easy.

Other random Toeplitz matrices have been studied recently in the probability literature. Bai (1999) posed the question of the existence of the LSD for $T_{n,\varepsilon} = ((\varepsilon_{|i-j|}))$ where $\{\varepsilon_t\}$ is i.i.d. with mean zero variance 1. Bryc, Dembo and Jiang (2006) and Hammond and Miller (2005) showed that then the LSD exists and is universal (does not depend on the underlying distribution of ε_1). Bose and Sen (2008) showed that the LSD of the Toeplitz matrix $T_{n,X} = ((X_{|i-j|}))$ exists when X satisfies (1.7) with some additional assumptions.

However, none of the above two results for random Toeplitz matrices are applicable to $\Gamma_n(X)$ due to the nonlinear dependence of $\hat{\gamma}_X(k)$ on $\{X_t\}$. We prove that under reasonable conditions on $\{\theta_k\}$ and $\{\varepsilon_t\}$ the LSD of $\Gamma_n(X)$ exists (see Theorem 1). In particular, this LSD is *universal* when $\{\varepsilon_t\}$ have mean zero and variance 1, are independent and, are either uniformly bounded or identically distributed. Further, in sharp contrast to the LSD of the S matrix, it has unbounded support. Moreover, it does not coincide with the LSD of $\Sigma_n(X)$.

Note that the LSD of $\Sigma_n(X)$ depends on the parameters $\{\theta_k\}$ but there is no one to one correspondence between $\{\theta_k\}$ and the LSD. For example the LSD is same when X is AR(1) with parameter θ or $-\theta$. The same situation persists for the LSD of $\Gamma_n(X)$ (see Theorem 5).

Incidentally, the only properties known for the LSD of $T_{n,\varepsilon}$ are that it is symmetric and has unbounded support. The moments of the LSD of $T_{n,X}$ when X_t is as in (1.7), may be written in a nice form involving $\{\theta_k\}$ and the moments of the LSD of $T_{n,\varepsilon}$. Unfortunately, a similar expression eludes us for the LSD of $\Gamma_n(X)$, primarily due to the nonlinear dependence of the autocovariances $\{\hat{\gamma}(k)\}$ on the driving $\{\epsilon_t\}$. We are thus unable to provide any explicit or implicit description of the LSD.

When $\{X_t\}$ is a finite order linear process, the limit moments in our case can be written as multinomial type sums of the autocovariances (see expression (2.4)). When X is of infinite order, the limit moments are the limiting values of these multinomial expressions as the order tends to infinity. Some additional properties of the limit moments are developed in Section 4. Apart from providing more information on the nature of the limit, some of these results are used crucially in the proof of Theorem 5.

The matrix $\Gamma_n(X)$ can be modified appropriately to rectify the facts that the matrices $\Gamma_n(X)$ and $\Sigma_n(X)$ have different LSD. This is based on the well known idea of *kernel density estimate.* For a sequence of integers $m = m_n \to \infty$, and a kernel function $K(\cdot)$ define

$$\hat{f}_X(t) = \sum_{k=-m}^{m} K(k/m) \exp(-2\pi i t k)\hat{\gamma}_X(k), \ t \in (0,\ 1]. \tag{1.8}$$

as the kernel density estimate of $f_X(\cdot)$. It is known that under suitable conditions $\hat{f}_X$ is a pointwise almost surely consistent estimate of f_X. Considering this as a spectral density, the corresponding autocovariance function is given by:

$$\begin{aligned}
\gamma_K(h) &= \int_{(0,\ 1]} \exp(2\pi i h x)\hat{f}_X(x)dx \\
&= \sum_{k=-m}^{m} K(k/m) \int_{(0,\ 1]} \exp\{2\pi i h x - 2\pi i x k\}\hat{\gamma}_X(k)dx \\
&= K(j/m)\hat{\gamma}_X(j) \ \text{ for all } \ -m \le j \le m.
\end{aligned}$$

This motivates the consideration of the *tapered* sample autocovariance matrix

$$\Gamma_{n,K}(X) = ((K((i-j)/m)\hat{\gamma}_X(i-j)))_{1 \le i,j \le n}. \tag{1.9}$$

It may be noted that if K is a nonnegative definite function then $\Gamma_{n,K}(X)$ is also nonnegative definite. Otherwise it may not be nonnegative definite.

Among other results, Xiao and Wu (2011) also showed that under the growth condition $m_n = o(n^\gamma)$ for a suitable γ and suitable conditions on K, the largest eigenvalue value of $\Gamma_{n,K}(X) - \Sigma_n(X)$ tends to zero. We show that (see Theorem 4) under the minimal condition $m/n \to 0$, if K is bounded, symmetric and continuous at 0 and $K(0) = 1$, then the LSD of $\Gamma_{n,K}(X)$ is indeed $f_X(U)$.

To deal with the inconsistency of the sample autocovariance matrix, the other idea in time series literature is to use banding. McMurray and Politis (2010) use such banded matrices while developing their bootstrap procedures. We study two such banded matrices. Let $\{m_n\}_{n\in\mathbb{N}} \to \infty$ be such that $\alpha_n := m_n/n \to \alpha \in [0,1]$. Then the *Type I banded sample autocovariance matrix* $\Gamma_n^{\alpha,I}(X)$ is same as $\Gamma_n(X)$ except that we substitute 0 for $\hat{\gamma}_X(k)$ whenever $k \geq m_n$. This is the same as the matrix $\Gamma_{n,K}$ with the choice $K(x) = I_{\{|x|\leq 1\}}$. The *Type II Banded Autocovariance Matrix* $\Gamma_n^{\alpha,II}(X)$ is the $m_n \times m_n$ principal sub matrix of $\Gamma_n(X)$. Theorem 3 states our results on these *banded autocovariance matrices*. In particular, the LSD exists for all α and is unbounded when $\alpha \neq 0$. When $\alpha = 0$, the LSD is $f_X(U)$.

Finally, a related matrix is,

$$\Gamma_n^*(X) = ((\gamma_X^*(|i-j|)))_{1\leq i,j\leq n} \text{ where } \gamma_X^*(k) = n^{-1}\sum_{i=1}^{n} X_i X_{i+k} \ , \ k = 0,1,\ldots. \tag{1.10}$$

Note that $\Gamma_n^*(X)$ is not nonnegative definite. This implies that many of the techniques applied to $\Gamma_n(X)$ are not available for $\Gamma_n^*(X)$. However, we are able to show that its LSD also exists but under stricter conditions on $\{X_t\}$ (see Theorem 2). Interestingly, simulations show that this LSD has significant positive mass on the negative axis. However, its moments dominate those of the LSD of $\Gamma_n(X)$ when $\theta_i \geq 0$ for all i (see Theorem 2(c)).

To illustrate our results, we provide a few simulation results for different choices of $\{\theta_k\}$. It would be nice to obtain additional theoretical properties of the ESD and the LSD of these matrices. For instance, the distribution of maximum eigenvalue of S has been studied in the literature. However, it does not seem to be at all easy to obtain similar results for $\Gamma_n(X)$.

Now a few words about the proofs. When $\alpha = 1$, then without loss of generality for asymptotic purposes, we assume that $m_n = n$. The full autocovariance matrix $\Gamma_n(X)$ may without loss of be visualised as a special case with $\alpha = 1$. The proof of Theorem 1(a) is quite long, primarily due to the nonlinear dependency of $\hat{\gamma}_X(\cdot)$ on $\{\epsilon_t\}$. See Section 3.3 for an outline description of the steps involved. In a nutshell, when $\{X_t\}$ is a finite order moving average process with bounded $\{\varepsilon_t\}$, we use the *method of moments* to establish the result. The assumption of boundedness is removed by the use of the *bounded Lipschitz metric* of convergence. We deal with the general case of infinite order by another use of this metric. Easy modifications of these arguments yield the existence of the LSD when $0 < \alpha \leq 1$. As we shall see later on, the case of $\alpha = 0$ is argued in a similar way. The proof of Theorem 4 is based on the arguments used in the proof of Theorem 1 and the ideas developed in Basak and Bose (2010) in the context of the study of weighted Toeplitz and Hankel matrices. The proof of Theorem 2 is a byproduct of the arguments in the proof of Theorem 1. However, due to the matrix now not being nonnegative definite, we impose the restriction that the random variables $\{\varepsilon_t\}$ are uniformly bounded.

2 Main results

We shall assume that $X = \{X_t\}_{t\in\mathbb{Z}}$ is a linear process (moving average process of possibly infinite order)

$$X_t = \sum_{k=0}^{\infty} \theta_k \varepsilon_{t-k} \tag{2.1}$$

where $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ is a sequence of independent random variables. A special case of this process is the so called MA(d) where $\theta_k = 0$ for all $k > d$. We denote this process by

$$X^{(d)} = \{X_{t,d} \equiv \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_d\varepsilon_{t-d}, t \in \mathbb{Z}\}$$

where without loss we assume that $\theta_0 \neq 0$.

It may also be mentioned that working with two sided moving average entails no difference. The conditions on $\{\varepsilon_t\}$ and on $\{\theta_k\}$ that will be used are:

Assumption A.

(a) $\{\varepsilon_t\}$ are *i.i.d.* with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t^2] = 1$.
(b) $\{\varepsilon_t\}$ are independent, uniformly bounded with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t^2] = 1$.

Assumption B.

(a) $\theta_j \geq 0$ for all j.

(b) $\sum_{j=0}^{\infty} |\theta_j| < \infty$.

It may be noted that the series in (2.1) converges almost surely under Assumption A(a) (or A(b)) and Assumption B(b). Further, X and $X^{(d)}$ are strongly stationary and ergodic under Assumption A(a) and weakly (second order) stationary under Assumption A(b) and Assumption B(b).

The autocovariance of X and $X^{(d)}$ are given by

$$\gamma_{X^{(d)}}(j) = \sum_{k=0}^{d-j} \theta_k\theta_{j+k} \quad \text{and} \quad \gamma_X(j) = \sum_{k=0}^{\infty} \theta_k\theta_{j+k}. \tag{2.2}$$

We now state our main results. We shall use the following notation: let $\{k_i\}$ stand for suitable integers.

$$\mathbf{k} = (k_0 \dots k_d), \quad S_{h,d} = \{\mathbf{k} : k_0, \dots, k_d \geq 0, k_0 + \cdots + k_d = h\}. \tag{2.3}$$

Theorem 1. *(Sample autocovariance matrix) Suppose Assumption A(a) or A(b) holds.*

(a) Then almost surely, $F^{\Gamma_n(X^{(d)})} \xrightarrow{w} F_d$ *which is nonrandom and does not depend on the distribution of* $\{\varepsilon_t\}$*. Further, for some sequence of constants* $\{p_{\mathbf{k}}^{(d)}\}$,

$$\beta_{h,d} = \int x^h dF_d(x) = \sum_{S_{h,d}} p_{\mathbf{k}}^{(d)} \prod_{i=0}^{d} [\gamma_{X^{(d)}}(i)]^{k_i}. \tag{2.4}$$

(b) Under Assumption B(b), almost surely, $F^{\Gamma_n(X)} \xrightarrow{w} F$ *which is nonrandom and independent of the distribution of* $\{\varepsilon_t\}$*. Further as* $d \to \infty$*,*

$$F_d \xrightarrow{w} F \quad \textit{and} \quad \beta_{h,d} \to \beta_h = \int x^h \, dF(x).$$

(c) Under Assumption B(a), F_d *has unbounded support and* $\beta_{h,d-1} \leq \beta_{h,d}$ *if* $d \geq 1$*. As a consequence, if Assumption B(a) and B(b) hold, then* F *has unbounded support.*

We now state an LSD result for $\Gamma_n^*(X)$. As mentioned before, this matrix is not nonnegative definite and this creates technical difficulties. We deal with only the case when Assumption A(b) holds and for simplicity we further assume that $\alpha = 1$.

Theorem 2. *Suppose Assumption A(b) holds.*

(a) Then almost surely, $F^{\Gamma_n^*(X^{(d)})} \xrightarrow{w} F_d^*$ *which is nonrandom and does not depend on the distribution of* $\{\varepsilon_t\}$*. For some constants* $\{p_{\mathbf{k}}^{*(d)}\}$*,*

$$\beta_{h,d}^* = \int x^h dF_d^*(x) = \sum_{S_{h,d}} p_{\mathbf{k}}^{*(d)} \prod_{i=0}^{d} [\gamma_{X^{(d)}}(i)]^{k_i}. \tag{2.5}$$

(b) Under Assumption B(b), almost surely $F^{\Gamma_n^*(X)} \xrightarrow{w} F^*$ *which is also nonrandom and does not depend on the distribution of* $\{\varepsilon_t\}$*. Further as* $d \to \infty$*,*

$$F_d^* \xrightarrow{w} F^* \quad \textit{and} \quad \beta_{h,d}^* \to \beta_h^* = \int x^h \, dF^*(x).$$

(c) Under Assumption B(a), F_d^* *has unbounded support and* $\beta_{h,d-1}^* \leq \beta_{h,d}^*$ *and* $\beta_{h,d} \leq \beta_{h,d}^*$*. Under Assumption B(a) and B(b),* F^* *has unbounded support. Moreover* $\beta_h \leq \beta_h^*$ *for all* h*.*

Remark 1.

(i) Observe that the expressions for moments in (2.4) and (2.5) are similar to the moments of $f_X(U)$ *given in 1.4. However, while the former two variables have unbounded support the latter has support contained in* $[-\sum_{-\infty}^{\infty} |\gamma_X(k)|, \ \sum_{-\infty}^{\infty} |\gamma_X(k)|]$*.*

(ii) Simulations show that the LSD of $\Gamma_n^*(X)$ *has positive mass on the negative real axis. Even then,* $\beta_h \leq \beta_h^*$ *for all* h*.*

(iii) Incidentally, the above results are in sharp contrast to the LSD of the S *matrix whose LSD is supported on the interval* $[0, 4]$*. See Bai (1999). See Remark 7 at the end of the proofs for a discussion on the assumptions required in different parts of the above two theorems. It turns out that the proof of the above theorem for* $d = \infty$ *is different from the proof of the previous theorem. This is because since there is no nonnegative definiteness, the bounded Lipschitz argument of Lemma 1 (b) cannot be used. The Assumption of all finite moments is also needed for the same reason.*

(iv) In the course of the proof of the above theorems, we shall show that $\{p_{\mathbf{k}}^{(d)}\}$ *of Theorem 1 satisfies*

$$p_{\mathbf{k}}^{(d)} \leq \frac{4^h (2h)!}{h!} \frac{h!}{k_0! \dots k_d!}.$$

As a consequence,

$$|\beta_{h,d}| \leq \frac{4^h(2h)!}{h!}(\sum_{k=0}^{d}|\theta_k|)^{2h} \quad \text{and} \quad |\beta_h| \leq \frac{4^h(2h)!}{h!}(\sum_{k=0}^{\infty}|\theta_k|)^{2h}$$

which are the even moments of a Gaussian random variable. Hence the limits have subexponential tails. The same is true for the LSD of $\Gamma_n^*(X)$.

Theorem 3. *(Banded sample autocovariance matrix) Suppose Assumption* $A(b)$ *holds.*

(a) Let $0 < \alpha \leq 1$. *Then all the conclusions of Theorem 1 hold for* $\Gamma_n^{\alpha,I}(X^{(d)})$ *and* $\Gamma_n^{\alpha,II}(X^{(d)})$ *with some modified constants* $\{p_{\mathbf{k}}^{\alpha,I,(d)}\}$ *and* $\{p_{\mathbf{k}}^{\alpha,II,(d)}\}$ *respectively in (2.4). Same conclusions continue to hold also for* $d = \infty$.

(b) When $\alpha = 0$, *and Assumption B(b) holds, the LSD of* $\Gamma_n^{\alpha,I}(X)$ *and* $\Gamma_n^{\alpha,II}(X)$ *are* $f_X(U)$.

In addition all the above remains true for $\Gamma_n^{\alpha,II}(X^{(d)})$ *and* $\Gamma_n^{\alpha,II}(X)$ *under the relaxed Assumption* $A(a)$.

Theorem 4. *(Tapered sample autocovariance matrix) Suppose Assumption A(a) or A(b) holds and Assumption B(b) holds. If* K *is bounded, symmetric and continuous at 0,* $K(0) = 1$, $K(x) = 0$ *for* $|x| > 1$ *and* $m_n \to \infty$ *such that* $m_n/n \to 0$. *Then the LSD of* $\Gamma_{n,K}(X)$ *is the same as that of* $\Sigma_n(X)$ *which equals* $f_X(U)$ *for* $d \leq \infty$.

Remark 2.

(i) If K *is nonnegative definite, then the theorem holds under Assumption A.*

(ii) Xiao and Wu (2011) show that under the assumption $m_n = o(n^\gamma)$ *(for a suitable* γ*) and other conditions, the maximum eigenvalue of* $\Sigma_n(X) - \Gamma_n(X)$ *tends to zero.*

The following result shows that different values of $\{\theta_k\}$ may give rise to the same LSD.

Theorem 5. *Under the conditions of the above Theorems, the LSD of* $\Gamma_n(X^{(d)})$ *(or* $\Gamma_n(X)$ *as the case maybe) are identical for the combinations* $(\theta_0, \theta_1, \theta_2, \ldots)$, $(\theta_0, -\theta_1, \theta_2, \ldots)$ *and* $(-\theta_0, \theta_1, -\theta_2, \ldots)$. *The result continues to hold for* $\Gamma_n^*(X)$.

Remark 3. *From the proof of Theorem 5 it will also be evident that the LSD of* $\Gamma_n(X^{(d)})$ *are identical for the processes which have autocovariances* $(\gamma_0, \gamma_1, \ldots, \gamma_d)$ *and* $(\gamma_0, -\gamma_1, \ldots, (-1)^d\gamma_d)$. *Same remark holds for Theorems 3 and 4. It may be noted that the LSD* $f_X(U)$ *of* $\Sigma_n(X)$ *has the same property.*

The following Figures 1 and 2 show the result of some simulation from the AR(1) and AR(2) models respectively.

3 Proofs

This section is structured as follows: in Section 3.1 we outline the so called Moment Method of establishing an LSD. This method is widely used in the random matrix literature and which

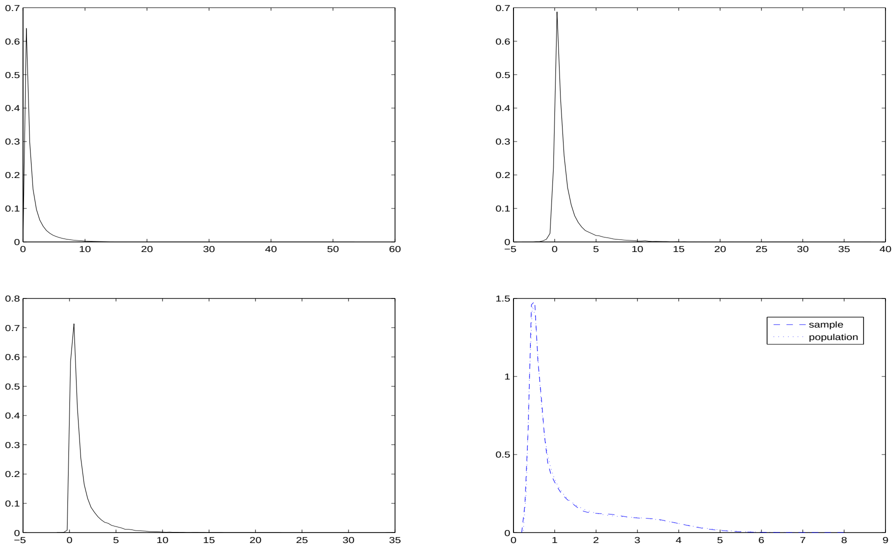

Figure 1: Kernel density estimates (KDEs) of the ESD, $n = 1000$, 100 realizations of: $\Gamma_n(X)$, $\alpha = 1$ (`top left`), $\alpha = 1/2$ (`bottom left`) and $\Gamma_n^*(X)$ (`top right`); $\Gamma_n(X)$, $\alpha \approx 0$, $(m = 10)$ (`dashed line`) and $\Sigma_n(X)$ (`dotted line`) `bottom right`. The input sequence is $X \sim AR(1)$, $\varepsilon_t \sim N(0, 1)$, $\theta = 1/2$.

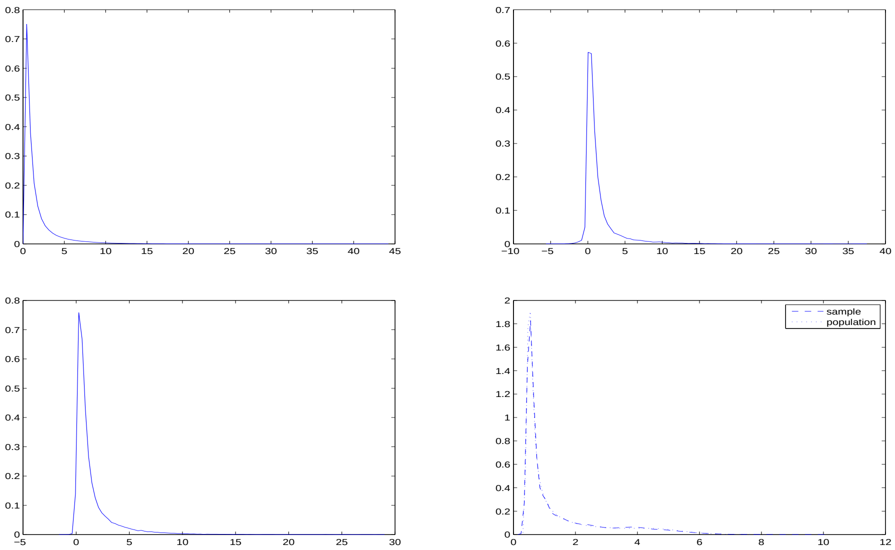

Figure 2: KDE of the ESD, $n = 1000$, 100 realizations of: $\Gamma_n(X)$, $\alpha = 1$ (`top left`), $\alpha = 1/2$ (`bottom left`) and $\Gamma_n^*(X)$ (`top right`); $\Gamma_n(X)$, $\alpha \approx 0$ $m = 10$ (`dashed line`) and of $\Sigma_n(X)$ (`dotted line`) `bottom right`. The input sequence is $X \sim AR(2)$, $\varepsilon_t \sim N(0, 1)$, $\theta_1 = 1/2$, $\theta_2 = 1/16$.

entails verifying three conditions which we label as (C1), (C2) and (C3). In Section 3.2 we introduce the well known bounded Lipschitz metric that metrizes the weak convergence of distribution functions and which will be crucially used in the proofs. In Section 3.3 we provide the proof of Theorem 1 (a). This proof is split up into the following parts: in Section 3.3.1 we show how to reduce the unbounded case of $\{\varepsilon_t\}$ to the bounded case by using the bounded Lipschitz metric; in Section 3.3.2 we develop a manageable expression for the moments of the empirical spectral distribution with a view to using the moment method; in Section 3.3.3 we show that only "matched" terms remain in the empirical moments from the asymptotic point of view. These moments are then written as an iterated sum, where one summation is over finitely many terms (each of which is called a "word"). Then in Section 3.3.4 we verify the crucial condition (C1) by showing that each one of these finitely many terms has a limit. This is the longest and hardest part of the proof. In Section 3.3.5 we verify (C2). In Section 3.3.6 we verify Carleman's condition (C3) and that finishes the proof of Theorem 1 (a). In Section 3.4 we provide the proof of Theorem 1 (b). In Section 3.5 we provide a proof of Theorem 1 (c) of the moment ordering (Section 3.5.1) and of the unbounded support (Section 3.5.2). In Section 3.6, we provide an outline of the proof of Theorem 3. In Section 3.8 we outline the proof of Theorem 2. Section 4 states and proves some properties of the limit moments. Recall that the limit moments equal $\sum p_{\mathbf{k}}^{(d)}$. In Lemma 6 and Lemma 7 several sufficient conditions are given for $p_{\mathbf{k}}^{(d)}$ to equal 0. Lemma 6 is used in the proof of Theorem 3. Lemma 7 provides a result which is parallel to a similar result proved by Bryc Dembo and Jiang (2006) that had turned out to be a crucial element in the proof of their main theorem. Lemma 9 shows that $p_{\mathbf{k}}^{(d)} = 0$ when $k_1 + k_3 + \cdots = \text{odd}$ and this is used in the proof of Theorem 5. This theorem is proved in Section 4.1.

3.1 Moment method

The moment method which may be described in brief as follows. Suppose $\{A_n\}$ is a sequence of $n \times n$ symmetric random matrices. Let $\beta_h(A_n)$ be the h^{th} moment of its ESD. It has the following nice form:

$$\beta_h(A_n) = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^h = \frac{1}{n}\operatorname{Tr}(A_n^h).$$

Suppose

(C1) $\mathbb{E}[\beta_h(A_n)] \longrightarrow \beta_h$ for all h (convergence of the average ESD).

(C2) $\sum_{n=1}^{\infty} \mathbb{E}\left[\beta_h(A_n) - \mathbb{E}[\beta_h(A_n)]\right]^4 < \infty$.

(C3) $\{\beta_h\}$ satisfies Carleman's condition: $\displaystyle\sum_{h=1}^{\infty} \beta_{2h}^{-1/2h} = \infty$.

Then LSD of $\{A_n\}$ exists almost surely and the limit distribution is uniquely identified by its moments $\{\beta_h\}$.

3.2 Bounded Lipschitz metric, d_{BL}

The *bounded Lipschitz metric* d_{BL}, is a complete metric defined on the space of probability measures on any Polish space (X, d), topologising the weak convergence of probability measures (see Dudley (2002)):

$$d_{BL}(\mu,\ \nu) = \sup\{\int f d\mu - \int f d\nu : ||f||_\infty + ||f||_L \leq 1\}$$

where

$$||f||_\infty = \sup_x |f(x)|, \quad ||f||_L = \sup_{x \neq y} |f(x) - f(y)|/d(x, y).$$

This metric will be used to estimate the distance between spectral measures via the following Lemma. Its proof may be found in Bai and Silverstein (2006) or Bai (1999) and uses Lidskii's theorem (see Bhatia, 1997, page 69).

Lemma 1. *(a) Suppose A and B are $n \times n$ real symmetric matrices. Then*

$$d_{BL}^2(F^A, F^B) \leq \frac{1}{n} \operatorname{Tr}(A - B)^2. \tag{3.1}$$

(b) Suppose A and B are $p \times n$ real matrices. Let $X = AA^T$ and $Y = BB^T$. Then

$$d_{BL}^2(F^X, F^Y) \leq \frac{2}{p^2} \operatorname{Tr}(X + Y) \operatorname{Tr}[(A - B)(A - B)^T]. \tag{3.2}$$

3.3 Proof of Theorem 1 (a)

A detailed proof will be provided only of Theorem 1 (a). We also note that if $\alpha = 1$ without loss of generality we can take $m_n = n$. The proof for $\alpha \in (0, 1)$ is quite similar to proof of Theorem 1 and hence we will provide a brief outline for that. The case $\alpha = 0$ will be a bit different and therefore we will give a somewhat detailed proof of that.

3.3.1 Reduction to bounded case using the metric d_{BL}

To use the moment method we need all moments to be finite. So the first step will be to show that we may without loss of generality, assume that $\{\varepsilon_t\}$ are uniformly bounded. For convenience, we will write

$$\Gamma_n(X^{(d)}) = \Gamma_{n,d}.$$

Lemma 1. *If for every $\{\varepsilon_t\}$ satisfying Assumption A(b), $\Gamma_n(X^{(d)})$ has the same LSD almost surely, then the same LSD continues to hold if $\{\varepsilon_t\}$ satisfies Assumption A(a).*

Proof. Suppose that for every $\{\varepsilon_t\}$ satisfying Assumption A(b), $\Gamma_n(X^{(d)})$ has the same LSD almost surely. Now suppose that $\{\varepsilon_t\}$ satisfies Assumption A(a). Define the bounded variables:

$$\widetilde{\varepsilon}_t = \varepsilon_t \mathbb{I}_{|\varepsilon_t| \leq C}, \qquad \widetilde{X}_{t,d} = \theta_0 \widetilde{\varepsilon}_t + \theta_1 \widetilde{\varepsilon}_{t-1} + \cdots + \theta_d \widetilde{\varepsilon}_{t-d},$$

$$\widehat{\varepsilon}_t = \frac{\widetilde{\varepsilon}_t - \mathbb{E}[\widetilde{\varepsilon}_t]}{\sqrt{\mathbf{Var}(\widetilde{\varepsilon}_t)}} \quad \text{and} \quad \widehat{X}_{t,d} = \theta_0\widehat{\varepsilon}_t + \theta_1\widehat{\varepsilon}_{t-1} + \cdots + \theta_d\widehat{\varepsilon}_{t-d}.$$

Let $\widetilde{\Gamma}_{n,d}$ and $\widehat{\Gamma}_{n,d}$ be the sample autocovariance matrices corresponding to $\widetilde{X}_{t,d}$ and $\widehat{X}_{t,d}$.

Let

$$A_{n,d} = \frac{1}{\sqrt{n}} \begin{bmatrix} 0 & X_{1,d} & X_{2,d} & \ldots & X_{n-1,d} & X_{n,d} & 0 & \ldots & 0 \\ 0 & 0 & X_{1,d} & \ldots & X_{n-2,d} & X_{n-1,d} & X_{n,d} & \ldots & 0 \\ & & & & \vdots & & & & \\ 0 & 0 & 0 & \ldots & 0 & X_{1,d} & X_{2,d} & \ldots & X_{n,d} \end{bmatrix}_{n\times 2n}$$

so that

$$\begin{aligned}(A_{n,d})_{i,j} &= X_{j-i,d}, \text{ if } 1 \leq j-i \leq n \\ &= 0, \text{ othewise.}\end{aligned}$$

Define $\widetilde{A}_{n,d}$ and $\widehat{A}_{n,d}$ in a similar way, using $\widetilde{X}_{t,d}$ and $\widehat{X}_{t,d}$.

Then

$$\Gamma_{n,d} = A_{n,d}A_{n,d}^T, \ \widetilde{\Gamma}_{n,d} = \widetilde{A}_{n,d}\widetilde{A}_{n,d}^T \ \text{ and } \ \widehat{\Gamma}_{n,d} = \widehat{A}_{n,d}\widehat{A}_{n,d}^T.$$

Note that by Lemma 1 (b),

$$\begin{aligned} d_{BL}^2(F^{\Gamma_{n,d}}, F^{\widehat{\Gamma}_{n,d}}) &\leq 2d_{BL}^2(F^{\Gamma_{n,d}}, F^{\widetilde{\Gamma}_{n,d}}) + 2d_{BL}^2(F^{\widetilde{\Gamma}_{n,d}}, F^{\widehat{\Gamma}_{n,d}}), \quad \text{and} \\ d_{BL}^2(F^{\Gamma_{n,d}}, F^{\widetilde{\Gamma}_{n,d}}) &\leq \frac{2}{n^2}\operatorname{Tr}[\Gamma_{n,d} + \widetilde{\Gamma}_{n,d}]\operatorname{Tr}[(A_{n,d} - \widetilde{A}_{n,d})(A_{n,d} - \widetilde{A}_{n,d})^T]. \end{aligned}$$

Now

$$\begin{aligned} \frac{1}{n}\operatorname{Tr}[\Gamma_{n,d} + \widetilde{\Gamma}_{n,d}] &= \frac{1}{n}\Big(\sum_{t=1}^{n} X_{t,d}^2 + \sum_{t=1}^{n} \widetilde{X}_{t,d}^2\Big) && (3.3) \\ &\leq \frac{1}{n}(1+d)\Big[\sum_{t=1}^{n}\sum_{k=0}^{d}\theta_k^2\varepsilon_{t-k}^2 + \sum_{t=1}^{n}\sum_{k=0}^{d}\theta_k^2\widetilde{\varepsilon}_{t-k}^2\Big] && (3.4) \\ &\leq \frac{2(1+d)}{n}\sum_{t=1}^{n}\sum_{k=0}^{d}\theta_k^2\varepsilon_{t-k}^2 \xrightarrow{a.e.} 2(1+d)\sum_{k=0}^{d}\theta_k^2 && (3.5) \end{aligned}$$

and

$$\begin{aligned} \frac{1}{n}\operatorname{Tr}[(A_{n,d} - \widetilde{A}_{n,d})(A_{n,d} - \widetilde{A}_{n,d})^T] &= \frac{1}{n}\sum_{t=1}^{n}(X_{t,d} - \widetilde{X}_{t,d})^2 \\ &\leq \frac{1+d}{n}\sum_{t=1}^{n}\sum_{k=0}^{d}\theta_k^2(\varepsilon_{t-k} - \widetilde{\varepsilon}_{t-k})^2 \\ &\xrightarrow{a.e.} (1+d)(\sum_{k=0}^{d}\theta_k^2)\mathbb{E}[\varepsilon_1^2\mathbb{I}_{|\varepsilon_1|\geq C}]. \end{aligned}$$

Hence

$$\limsup_{n\to\infty} d_{BL}^2(F^{\Gamma_{n,d}}, F^{\widetilde{\Gamma}_{n,d}}) \leq 4(1+d)^2 \left(\sum_{k=0}^{d} \theta_k^2\right)^2 \mathbb{E}[\varepsilon_1^2 \mathbb{I}_{|\varepsilon_1|\geq C}] \ a.s. \tag{3.6}$$

Similarly,

$$\limsup_{n\to\infty} d_{BL}^2(F^{\widetilde{\Gamma}_{n,d}}, F^{\widehat{\Gamma}_{n,d}}) \leq k \left[\left(1 - \frac{1}{\sigma(C)}\right)^2 + \frac{\mu^2(C)}{\sigma^2(C)}\right] \ a.s. \tag{3.7}$$

where k is a constant, $\mu(C) = \mathbb{E}[\varepsilon_1 \mathbb{I}_{|\varepsilon_1|>C}]$ and $\sigma^2(C) = \mathbf{Var}(\widetilde{\varepsilon}_1)$.

From the hypothesis, LSD of $F^{\widehat{\Gamma}_{n,d}}$ exists and is free of C. On the other hand, as $C \to \infty$, $\mu(C) \to 0$ and $\sigma(C) \to 1$. It follows from (3.6) and (3.7) that

$$\lim_{C\to\infty} \limsup_{n\to\infty} d_{BL}^2(F^{\Gamma_{n,d}}, F^{\widehat{\Gamma}_{n,d}}) = 0 \ \text{ a.s..}$$

The Lemma then follows immediately. □

Thus from now on we assume that Assumption A(b) holds.

3.3.2 Manageable expression for $\beta_h(\Gamma_{n,d})$

Recall from Section 3.1 condition (C1), that we need to prove the convergence for every moment. Thus, fix any arbitrary positive integer h and consider the h^{th} moment. We note that,

$$\Gamma_{n,d} = \frac{1}{n}((Y_{i,j}^{(n)}))_{i,j=1,\ldots,n}, \text{ where } Y_{i,j}^{(n)} = \sum_{t=1}^{n} X_{t,d} X_{t+|i-j|,d} \mathbb{I}_{(t+|i-j|\leq n)}. \tag{3.8}$$

The h^{th} moment of the ESD of $\Gamma_{n,d}$ is

$$\begin{aligned}
\beta_h(\Gamma_{n,d}) &= \frac{1}{n}\operatorname{Tr}(\Gamma_{n,d}^h) \\
&= \frac{1}{n^{h+1}} \sum_{\substack{1\leq \pi_0,\ldots,\pi_h\leq n \\ \pi_h=\pi_0}} Y_{\pi_0,\pi_1}^{(n)} \cdots Y_{\pi_{h-1},\pi_h}^{(n)} \\
&= \frac{1}{n^{h+1}} \sum_{\substack{1\leq \pi_0,\ldots,\pi_h\leq n \\ \pi_h=\pi_0}} \Big[\prod_{j=1}^{h} \Big(\sum_{t_j=1}^{n} X_{t_j,d} X_{t_j+|\pi_j-\pi_{j-1}|,d} \mathbb{I}_{(t_j+|\pi_j-\pi_{j-1}|\leq n)}\Big)\Big]. \qquad (3.9)
\end{aligned}$$

To express the above in a neater and more amenable form, define

$$\begin{aligned}
\mathbf{t} &= (t_1,\ldots,t_h), \quad \pi = (\pi_0,\ldots,\pi_{h-1}), \\
\mathcal{A} &= \Big\{(\mathbf{t},\pi) : 1 \leq t_1,\ldots,t_h,\pi_0,\ldots,\pi_{h-1} \leq n, \pi_h = \pi_0\Big\}, \\
\mathbf{a}(\mathbf{t},\pi) &= (t_1,\ldots,t_h, t_1 + |\pi_0 - \pi_1|, \ldots, t_h + |\pi_{h-1} - \pi_h|), \\
\mathbf{a} &= (a_1,\ldots,a_{2h}) \in \{1,2,\ldots,2n\}^{2h}, \quad X_{\mathbf{a}} = \prod_{j=1}^{2h}(X_{a_j,d}) \text{ and } \mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)} = \prod_{j=1}^{h} \mathbb{I}_{(t_j+|\pi_{j-1}-\pi_j|\leq n)}.
\end{aligned}$$

Then using (3.9) we can write the so called *trace formula*,

$$\mathbb{E}[\beta_h(\Gamma_{n,d})] \;=\; \frac{1}{n^{h+1}}\mathbb{E}\Big[\sum_{(\mathbf{t},\pi)\in\mathcal{A}} X_{\mathbf{a}(\mathbf{t},\pi)}\mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)}\Big]. \tag{3.10}$$

3.3.3 Matching and negligibility of certain terms

Note that by independence of $\{\varepsilon_t\}$, $\mathbb{E}[X_{\mathbf{a}(\mathbf{t},\pi)}] = 0$ if there is at least one component of the product that has no ε_t common with any other component. Motivated by this, we introduce a notion of appropriate matching and then show that certain higher order terms can be asymptotically neglected in the trace formula (3.10).

We say

- $\mathbf{a}$ is **d-matched** (in short *matched*) if $\forall i \le 2h, \exists j \ne i$ such that $|a_i - a_j| \le d$. Note that when $d = 0$ this means $a_i = a_j$.
- $\mathbf{a}$ is **minimal d-matched** (in short minimal matched) if there is a partition $\mathcal{P}$ of $\{1,\ldots,2h\}$, such that

$$\{1,\ldots,2h\} = \cup_{k=1}^{h}\{i_k, j_k\},\; i_k < j_k$$

such that

$$|a_x - a_y| \le d \Leftrightarrow \{x,y\} = \{i_k, j_k\} \text{ for some } k.$$

For example, for $d = 1$, $h = 3$, $(1,2,3,4,9,10)$ is matched but not minimal matched and $(1,2,5,6,9,10)$ is both matched and minimal matched.

Lemma 2.

$$\#\{\mathbf{a} :\ \mathbf{a}\ \textit{is matched but not minimal matched}\ \} = O(n^{h-1}). \tag{3.11}$$

Proof. Consider the graph with $2h$ vertices $\{1,2,\ldots,2h\}$ where we join vertices i and j with an edge if $|a_i - a_j| \le d$. Let $k = \#$ connected components. Suppose $\mathbf{a}$ is matched but not minimal matched. Let

$$l_j = \#\ \text{ vertices in the } j\text{-th component.}$$

Since $\mathbf{a}$ is matched, $l_j \ge 2$ for all j and $l_j > 2$ for at least one j. Hence

$$2h = \sum_{j=1}^{k} l_j > 2k \text{ which implies } k \le h-1.$$

Also if i and j are in the same component then $|a_i - a_j| \le 2dh$. Hence for each connected component, there are $O(n)$ many choices for the a_i's for which i belongs to that component. Hence the result follows. □

Now we can rewrite (3.10) as

$$\mathbb{E}[\beta_h(\Gamma_{n,d})] \;=\; \frac{1}{n^{h+1}}\mathbb{E}[\sum_1 X_{\mathbf{a}(\mathbf{t},\pi)}\mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)}] + \frac{1}{n^{h+1}}\mathbb{E}[\sum_2 X_{\mathbf{a}(\mathbf{t},\pi)}\mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)}] + \frac{1}{n^{h+1}}\mathbb{E}[\sum_3 X_{\mathbf{a}(\mathbf{t},\pi)}\mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)}]$$

$$= \quad T_1 + T_2 + T_3 \quad \text{(say)}.$$

where $\sum_i$, $i = 1, 2, 3$ are summations taken over all $(\mathbf{t}, \pi) \in \mathcal{A}$ such that $\mathbf{a}(\mathbf{t}, \pi)$ is respectively, (i) minimal matched, (ii) matched but not minimal matched and (iii) not matched.

By mean zero assumption, $T_3 = 0$. Since X_i's are uniformly bounded, by using Lemma 2, $T_2 \le \frac{C}{n}$ for some constant C. So provided the limit exists,

$$\lim_{n\to\infty} \mathbb{E}[\beta_h(\Gamma_{n,d})] = \lim_{n\to\infty} \frac{1}{n^{h+1}} \mathbb{E}[\sum_{\substack{(\mathbf{t},\pi)\in\mathcal{A}:\mathbf{a}(\mathbf{t},\pi) \text{ is} \\ \text{minimal matched}}} X_{\mathbf{a}(\mathbf{t},\pi)} \mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)}]. \tag{3.12}$$

Hence, from now our focus will be only on minimal matched words.

3.3.4 Verification of (C1) for Theorem 1 (a)

As mentioned earlier, this is the hardest and lengthiest part of the proof. One can give a separate and easier proof for the case $d = 0$. However, the proof for general d and the simpler case of $d = 0$ are developed in parallel since this helps us to relate the limit for general d to limit for $d = 0$.

The idea behind the proof of (C1) is as follows. Our starting point is equation (3.12). We first define an *equivalence relation* (see below) on the set of minimal matched $\mathbf{a} = \mathbf{a}(\mathbf{t}, \pi)$ which gives rise to finitely many equivalence classes (see (3.14)). Using this, we write the sum in (3.12) as an iterated sum where the outer sum is over the finite number of equivalence classes (see (3.15)). Then we show that for every fixed equivalence class, the inner sum has a limit.

To define the equivalence relation, consider the collection of $(2d + 1)h$ symbols (letters)

$$\mathcal{W}_h = \{w^k_{-d}, \ldots, w^k_0, \ldots, w^k_d : \ k = 1, \ldots, h\}.$$

Suppose $\mathbf{a} = (a_1, \ldots, a_{2h})$ is minimal d matched. Then it induces the corresponding partition

$$\mathcal{P} = \cup_{k=1}^h \{i_k, j_k\} \quad \text{with} \quad i_k < j_k \quad \text{of} \quad \{1, \ldots, 2h\}$$

where the $\{i_k\}$ are arranged in *ascending order*. With this $\mathbf{a}$, associate the **word** $w = w[1]w[2] \ldots w[2h]$ of length $2h$ where

$$w[i_k] = w^k_0, \ w[j_k] = w^k_l \quad \text{if} \quad a_{i_k} - a_{j_k} = l, \ 1 \le k \le h. \tag{3.13}$$

As an example, consider $d = 1, h = 3$ and $\mathbf{a} = (a_1, \ldots, a_6) = (1, 21, 1, 20, 39, 40)$. Then the unique partition of $\{1, 2, \ldots, 6\}$ and the unique word associated with $\mathbf{a}$ are $\{\{1, 3\}, \{2, 4\}, \{5, 6\}\}$ and $[w^1_0 w^2_0 w^1_0 w^2_1 w^3_0 w^3_{-1}]$ respectively.

It is important to note that corresponding to any fixed partition $\mathcal{P} = \{\{i_k, j_k\}, 1 \le k \le h\}$, there are several $\mathbf{a}$ associated with it and there are exactly $(2d + 1)^h$ words that can arise from it. For example with $d = 1, h = 2$ consider the partition $\mathcal{P} = \{\{1, 2\}, \{3, 4\}\}$. Then the nine words corresponding to $\mathcal{P}$ are $w^1_0 w^1_i w^2_0 w^2_j$ where $i, j = -1, 0, 1$.

By a slight abuse of notation we write $w \in \mathcal{P}$ if the partition corresponding to w is same as $\mathcal{P}$. We will say that

• $w[x]$ *matches* with $w[y]$ (denote by $w[x] \approx w[y]$) iff $w[x] = w_l^k$ and $w[y] = w_{l'}^k$ for some k, l, l'.

• w is d *pair matched* if it is *induced* by a minimal d matched $\mathbf{a}$ (so $w[x]$ matches with $w[y]$ iff $|a_x - a_y| \leq d$).

Clearly this induces an *equivalence relation* on all d minimal matched $\mathbf{a}$ and the equivalence classes can be indexed by d pair matched w. Given a d pair matched w, the corresponding equivalence class is given by

$$\begin{aligned}\Pi(w) \quad = \quad & \{(\mathbf{t},\pi) \in \mathcal{A} : \ w[i_k] = w_0^k, w[j_k] = w_l^k \Leftrightarrow \\ & \mathbf{a}(\mathbf{t},\pi)_{i_k} - a(\mathbf{t},\pi)_{j_k} = l \ \text{ and } \ \mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)} = 1\}.\end{aligned} \tag{3.14}$$

Then we may rewrite (3.12) as (provided the second limit exists)

$$\lim_{n\to\infty} \mathbb{E}[\beta_h(\Gamma_{n,d})] = \sum_{\mathcal{P}} \sum_{w\in\mathcal{P}} \lim_{n\to\infty} \frac{1}{n^{h+1}} \sum_{(t,\pi)\in\Pi(w)} \mathbb{E}[X_{\mathbf{a}(\mathbf{t},\pi)} \mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)}]. \tag{3.15}$$

By using the autocovariance structure, we further simplify the above as follows. Let

$$\mathcal{W}(\mathbf{k}) = \{w : \ \#\{s : |w[i_s] - w[j_s]| = i\} = k_i, \ i = 0, 1, \ldots, d\}.$$

Using the definitions of $\gamma_{X^{(d)}}(\cdot)$ and of $S_{h,d}$ given in (2.3), we may rewrite (3.15) as

$$\lim_{n\to\infty} \mathbb{E}[\beta_h(\Gamma_{n,d})] = \sum_{\mathcal{P}} \sum_{S_{h,d}} \sum_{w\in\mathcal{P}\cap\mathcal{W}(\mathbf{k})} \lim_{n\to\infty} \frac{1}{n^{h+1}} |\Pi(w)| \prod_{i=0}^{d} [\gamma_{X^{(d)}}(i)]^{k_i} \tag{3.16}$$

provided

$$p_w^{(d)} \equiv \lim_{n\to\infty} \frac{1}{n^{h+1}} |\Pi(w)| \tag{3.17}$$

exists for every word w of length $2h$.

To show that this limit exists, it is convenient to work with $\Pi^*(w) \supseteq \Pi(w)$ defined as

$$\begin{aligned}\Pi^*(w) \quad = \quad & \{(\mathbf{t},\pi) \in \mathcal{A} : \ w[i_k] = w_0^k, \ w[j_k] = w_l^k \Rightarrow \\ & a(\mathbf{t},\pi)_{i_k} - a(\mathbf{t},\pi)_{j_k} = l \ \text{ and } \ \mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)} = 1\}.\end{aligned} \tag{3.18}$$

By Lemma 2 we have for every w,

$$\frac{1}{n^{h+1}} |\Pi^*(w) - \Pi(w)| \to 0.$$

Thus it is enough to show that $\lim_{n\to\infty} \frac{1}{n^{h+1}} |\Pi^*(w)|$ exists.

For a pair matched w we divide the coordinates of w according to the position of the matches as follows. Let

$$S_1(w) \quad = \quad \{i : \ w[i] \ \text{ matches with } \ w[j] \ \text{ for some } \ j, \ i < j \leq h\}$$

$$
\begin{aligned}
S_2(w) &= \{j : \ w[i] \ \text{ matches with } \ w[j] \ \text{ for some } \ i, \ i < j \leq h\} \\
S_3(w) &= \{i : \ w[i] \ \text{ matches with } \ w[j] \ \text{ for some } \ j, \ i \leq h < j\} \\
S_4(w) &= \{j : \ w[i] \ \text{ matches with } \ w[j] \ \text{ for some } \ i, \ i \leq h < j\} \\
S_5(w) &= \{i : \ w[i] \ \text{ matches with } \ w[j] \ \text{ for some } \ j, \ h < i < j\} \\
S_6(w) &= \{j : \ w[i] \ \text{ matches with } \ w[j] \ \text{ for some } \ i, \ h < i < j\}.
\end{aligned}
$$

Let

$$E = \{t_1, \ldots, t_h, \pi_0, \ldots, \pi_h\}.$$

Let

$$G = \{t_i | i \in S_1(w) \cup S_3(w)\} \cup \{\pi_0\} \cup \{\pi_i | i + h \in S_5(w)\}.$$

Note that $G \subset E$. Elements in G will be called the *generating vertices*. These are the indices where the first occurrence of any matched letter happens. Note that G has $(h+1)$ elements say $u_1^n, \ldots, u_{h+1}^n$ and for simplicity we will write

$$G \equiv U^n = (u_1^n, \ldots, u_{h+1}^n) \ \text{ and } \ \mathcal{N}_n = \{1, 2, \ldots, n\}.$$

Claim 1: Each element of E may be written as a linear expression, (say $\lambda_\mathbf{i}$) of the generating vertices that are all to the *left* of the element.

Proof of Claim 1: We shall denote the constants in the proposed linear expressions by $\{m_j\}$.

(a) For those elements of E that are generating vertices, we take the constants as $m_j = 0$ and the linear combination is taken as the identity mapping so that

$$
\begin{aligned}
\text{for all } \ i \in S_1(w) \cup S_3(w), \ \lambda_\mathbf{i} &\equiv t_i, \\
\lambda_\mathbf{h+1} &\equiv \pi_0, \\
\text{and for all } \ i + h \in S_5(w), \ \lambda_\mathbf{i+h+1} &\equiv \pi_i.
\end{aligned}
$$

(b) Using the relations between $S_1(w)$ and $S_2(w)$ induced by w, we can write

$$\text{for all } \quad j \in S_2(w), \ t_j = \lambda_\mathbf{j} + n_j,$$

for some n_j such that $|n_j| \leq d$ and define $m_j = n_j$ for $j \in S_2(w)$ and $\lambda_\mathbf{j} \equiv \lambda_\mathbf{i}$.

(c) It remains to write π_j's for $j + h \in S_4(w) \cup S_6(w)$.

Note that for every π we can write

$$|\pi_{i-1} - \pi_i| = b_i(\pi_{i-1} - \pi_i) \ \text{ for some } \ b_i \in \{-1, 1\}.$$

Consider the vector

$$\mathbf{b} = (b_1, b_2, \ldots, b_h) \in \{-1, 1\}^h.$$

Thus $\mathbf{b}$ will be a valid choice if we have

$$b_i(\pi_{i-1} - \pi_i) \geq 0 \ \text{ for all } \ i. \tag{3.19}$$

We then have the following two cases:

Case 1: $w[i]$ matches with $w[j+h]$, $j+h \in S_4(w)$ and $i \in S_3(w)$. Then we get

$$t_i = t_j + b_j(\pi_{j-1} - \pi_j) + n_{j+h} \text{ for some integer } n_{j+h} \in \{-d, \dots, 0 \dots, d\}. \quad (3.20)$$

Case 2: $w[i+h]$ matches with $w[j+h]$, $j+h \in S_6(w)$ and $i+h \in S_5(w)$. Then we have

$$t_i + |\pi_{i-1} - \pi_i| = t_j + |\pi_{j-1} - \pi_j| + n_{j+h} \text{ where } n_{j+h} \in \{-d, \dots, 0, \dots, d\}. \quad (3.21)$$

So we note that inductively from left to right we can write

$$\pi_j = \lambda^{\mathbf{b}}_{\mathbf{j+1+h}} + m_{j+1+h}, \ j+h \in S_4(w) \cup S_6(w). \quad (3.22)$$

Hence we can inductively write π_j as a linear combination $\{\lambda^{\mathbf{b}}_{\mathbf{j}}\}$ of the generating vertices up to some appropriate constant. Here we used the superscript $\mathbf{b}$ to emphasize the fact that the linear expression will depend on $\mathbf{b}$. Also note that by construction, the linear functions $\{\lambda^{\mathbf{b}}_{\mathbf{j}}\}$ for any element depends *only* on the vertices present to the left of it. This proves our claim. □

Now we are almost ready to write down an expression for the limit. If $\lambda_{\mathbf{i}}$ were unique for each $\mathbf{b}$, then we could write the $|\Pi^*(w)|$ as a sume of all possible choices of b and we could tackle the expression for each b separately. However, $\lambda_{\mathbf{i}}$'s may be same for several choices $b_i \in \{-1, 1\}$. For example, for the word $w_0^1 w_0^2 w_0^1 w_0^2$, we may choose any $\mathbf{b}$. We first circumvent this problem in the following way: Let

$$\mathcal{T} = \{j+h \in S_4(w) \cup S_6(w) \mid \ \lambda^{\mathbf{b}}_{\mathbf{j+h}} - \lambda^{\mathbf{b}}_{\mathbf{j+h-1}} \equiv 0 \ \ \forall b_j\}.$$

Note that the definition of $\mathcal{T}$ depends on w only through the partition $\mathcal{P}$ it generates.

Suppose $j+h \in \mathcal{T}$. Then from (3.20) and (3.21) the region given by (3.19) is

$$b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(U^n) - \lambda^{\mathbf{b}}_{\mathbf{j+h}}(U^n) + m_{j+h-1} - m_{j+h}) \geq 0.$$

Claim 2: The above expression is same for all choices of $\{b_j\}$.

Proof of Claim 2: Here is a short proof of the claim. First we show that if $j+h \in \mathcal{T}$ then we must have

$$t_j = t_j + |\pi_{j-1} - \pi_j| + n_j \text{ for some integer } |n_j| \leq d. \quad (3.23)$$

Suppose that this is not true.

So first assume that $j+h \in S_6(w)$. Then we will have a relation

$$t_i + b_i(\pi_{i-1} - \pi_i) = t_j + b_j(\pi_{j-1} - \pi_j) + n_j, \text{ where } i+h \in S_5(w).$$

Recall that any typical linear function, $\lambda^{\mathbf{b}}_{\mathbf{j}}$ depends only the vertices present to the left of it. Thus in the above equation coefficient of π_i would be nonzero and hence we must have $\lambda^{\mathbf{b}}_{\mathbf{j+h-1}} - \lambda^{\mathbf{b}}_{\mathbf{j+h}} \not\equiv 0$.

Now assume $j+h \in S_4(w)$ and $w[i]$ matches with $w[j+h]$ for $i \neq j$ then we can repeat the argument above to arrive at a similar contradiction. This shows that if $j+h \in \mathcal{T}$ then our relation must be like (3.23). Now a simple calculation shows that for relations like (3.23) we have

$$b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(U^n) - \lambda^{\mathbf{b}}_{\mathbf{j+h}}(U^n) + m_{j+h-1} - m_{j+h}) = -n_j$$

which is of course same across all choices of $\mathbf{b}$. This proves our claim. $\square$

Now note that if $j+h \in \mathcal{T}$ and if $n_{j+h} \neq 0$ then as we change b_j it does change the value of m_{2h+1}. Further, we can have at most two choices for π_j for every choices of π_{j-1} if $n_{j+h} \neq 0$ depending on b_j.

However for $j \in \mathcal{T}$ and $n_j = 0$ we have only one choice for π_j given the choice for π_{j-1} for every choice of b_j. On the other hand we know that $\mathbf{b} \in \{-1,1\}^h$ must satisfy (3.19).

Considering all this, let

$$\mathcal{B}(w) = \{\mathbf{b} \in \{-1,1\}^h \mid b_j = 1 \text{ if } n_j = 0 \text{ for } j \in \mathcal{T}\}$$

where $\{n_j\}$ is as in Claim 2. Then we may write,

$$\begin{aligned} p_w^{(d)} := \lim_n \frac{1}{n^{h+1}} |\Pi^*(w)| \;=\; & \lim_n \frac{1}{n^{h+1}} \sum_{\mathbf{b} \in \mathcal{B}(w)} \sum_{U^n \in \mathcal{N}_n^{h+1}} \mathbb{I}({\lambda^{\mathbf{b}}}_{2h+1}(U^n) + m_{2h+1} = {\lambda^{\mathbf{b}}}_{h+1}(U^n) + m_{h+1}) \\ & \times \prod_{j=1}^{2h} \mathbb{I}(\lambda_{\mathbf{j}}^{\mathbf{b}}(U^n) + m_j \in \mathcal{N}_n) \\ & \times \prod_{\substack{j=1 \\ j \notin \mathcal{T}}}^{h} \mathbb{I}(b_j(\lambda_{\mathbf{j+h-1}}^{\mathbf{b}}(U^n) - \lambda_{\mathbf{j+h}}^{\mathbf{b}}(U^n) + m_{j+h-1} - m_{j+h}) \geq 0) \\ & \times \prod_{j=1}^{h} \mathbb{I}(\lambda_{\mathbf{j}}^{\mathbf{b}}(U^n) + b_j(\lambda_{\mathbf{j+h-1}}^{\mathbf{b}}(U^n) - {\lambda^{\mathbf{b}}}_{j+h}(U^n) + m_{j+h-1} - m_{j+h}) \leq n) \\ & \times \prod_{j \in \mathcal{T}} \mathbb{I}(n_j \leq 0). \end{aligned} \tag{3.24}$$

Note that the right side of (3.24) can be rewritten as

$$\begin{aligned} \lim_n \sum_{\mathbf{b} \in \mathcal{B}(w)} \mathbb{P}_n \Big(& \lambda_{2h+1}^{\mathbf{b}}(U^n) + m_{2h+1} = \lambda_{\mathbf{h+1}}^{\mathbf{b}}(U^n) + m_{h+1}; \lambda_{\mathbf{j}}^{\mathbf{b}}(U^n) + m_j \in \mathcal{N}_n, 1 \leq j \leq 2h; \\ & b_j(\lambda_{\mathbf{j+h-1}}^{\mathbf{b}}(U^n) - \lambda_{\mathbf{j+h}}^{\mathbf{b}}(U^n) + m_{j+h-1} - m_{j+h}) \geq 0, 1 \leq j \leq h,\ j \notin \mathcal{T}; \\ & \lambda_{\mathbf{j}}^{\mathbf{b}}(U^n) + b_j(\lambda_{\mathbf{j+h-1}}^{\mathbf{b}}(U^n) - \lambda_{\mathbf{j+h}}^{\mathbf{b}}(U^n) + m_{j+h-1} - m_{j+h}) \leq n, 1 \leq j \leq h; \\ & n_j \leq 0, j \in \mathcal{T} \Big), \end{aligned}$$

where the probability is computed under the distribution $\mathbb{P}_n$ of discrete uniform on $\mathcal{N}_n^{h+1}$.

Fix a partition $\mathcal{P}$ and $\mathbf{b} \in \{-1,1\}^h$. If $d = 0$, then there is one and only one word corresponding to it. However, across any d and any fixed $k_0, k_1, \ldots, k_d$, the linear functions $\lambda_{\mathbf{j}}$'s continue to remain same. The only possible changes will be in the values of m_j's. *This is why there is a relation between the limit for* $d = 0$ *and* $d \neq 0$.

We now identify the cases where the above limit is zero.

Claim: Suppose w is such that $\mathcal{R} := \Big\{\lambda_{2h+1}^{\mathbf{b}}(U^n) + m_{2h+1} = \lambda_{\mathbf{h+1}}^{\mathbf{b}}(U^n) + m_{h+1}\Big\}$ is a lower dimensional subset of $\mathcal{N}_n^{h+1}$. Then the above limit is zero.

Proof: First consider the case $d = 0$. Then $m_j = 0, \forall j$. So Note that $\mathcal{R}$ lies in a hypercube. Hence the result follows by convergence of the Riemann sum to the corresponding Riemann integral.

For any general d the corresponding region is just a translate of the region considered for $= 0$. Hence the result follows. □

Hence for a fixed $w \in \mathcal{P}$, a positive limit contribution is possible only when $\mathcal{R} = \mathcal{N}_n^{h+1}$. This implies that we must have

$$\begin{aligned} \lambda^{\mathbf{b}}_{\mathbf{2h+1}}(U^n) - \lambda^{\mathbf{b}}_{\mathbf{h+1}}(U^n) &\equiv 0 \ \ (\text{for } \ d=0) \\ \lambda^{\mathbf{b}}_{\mathbf{2h+1}}(U^n) - \lambda^{\mathbf{b}}_{\mathbf{h+1}}(U^n) &\equiv 0 \ \text{ and } \ m_{2h+1} - m_{h+1} = 0 \ \ (\text{for general } \ d). \end{aligned}$$

Note that the first relation depends only the partition $\mathcal{P}$ but the second relation is determined by the word w.

Now note that $\lambda^{\mathbf{b}}_{\mathbf{j}}$ are linear forms with integer coefficients

$$\lambda^{\mathbf{b}}_{\mathbf{j}}(U^n) + m_j \in \{1, \ldots, n\} \iff \lambda^{\mathbf{b}}_{\mathbf{j}}\left(\frac{U^n}{n}\right) + \frac{m_j}{n} \in (0,1].$$

Noting that $\frac{U^n}{n} \stackrel{w}{\Rightarrow} U$ following uniform distribution on $[0,1]^{h+1}$, $\frac{1}{\lim} n^{h+1} |\Pi^*(w)| = p_w^{(d)}$ equals

$$\sum_{\mathbf{b} \in \mathcal{B}(w)} \mathbb{P}(\lambda^{\mathbf{b}}_{\mathbf{j}}(U) \in (0,1), \ 1 \le j \le 2h; \ b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(U) - \lambda^{\mathbf{b}}_{\mathbf{j+h}}(U)) \ge 0, \ 1 \le j \le h, j \notin \mathcal{T};$$
$$\lambda^{\mathbf{b}}_{\mathbf{j}}(U) + b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(U) - \lambda^{\mathbf{b}}_{\mathbf{j+h}}(U)) \le 1, 1 \le j \le h; \ \lambda^{\mathbf{b}}_{\mathbf{2h+1}}(U) = \lambda^{\mathbf{b}}_{\mathbf{h+1}}(U))$$
$$\times \mathbb{I}(m_{2h+1} = m_{h+1}) \times \mathbb{I}(n_j \le 0, j \in \mathcal{T}). \tag{3.25}$$

Let us denote

$$p_{\mathbf{k}}^{\mathcal{P},d} = \sum_{w \in \mathcal{P} \cap \mathcal{W}(\mathbf{k})} p_w^{(d)} \ \text{ and } \ p_{\mathbf{k}}^{(d)} = \sum_{\mathcal{P}} p_{\mathbf{k}}^{\mathcal{P},d}. \tag{3.26}$$

Thus the expression (3.16) becomes

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}[\beta_h(\Gamma_{n,d})] &= \sum_{\mathcal{P}} \sum_{\mathbf{k} \in S_{h,d}} p_{\mathbf{k}}^{\mathcal{P},d} \prod_{i=0}^{d} [\gamma_{X^{(d)}}(i)]^{k_i} & (3.27) \\ &= \sum_{\mathbf{k} \in S_{h,d}} p_{\mathbf{k}}^{(d)} \prod_{i=0}^{d} [\gamma_{X^{(d)}}(i)]^{k_i}. & (3.28) \end{aligned}$$

So we have proved convergence of expected moments of the ESDs of $\Gamma_{n,d}$. Thus verification of (C1) is complete. □

Remark 4. *From the above discussion, observing that the indicators in (3.26) are one when $d = 0$, it follows that*

$$p_w^{(d))} \le p_w^0 2^h.$$

As a consequence we have

$$p_{\mathbf{k}}^{\mathcal{P},d} \le 2^h \binom{h}{k_0, k_1, \ldots, k_d} p_h^{\mathcal{P},0} \tag{3.29}$$

since

3.3.5 Verification of (C2) for Theorem 1 (a)

We state this formally in the following Lemma.

Lemma 3.
$$\mathbb{E}\left[\frac{1}{n}\operatorname{Tr}(\Gamma^h_{n,d}) - \frac{1}{n}\mathbb{E}[\operatorname{Tr}(\Gamma^h_{n,d})\right]^4 = O\left(\frac{1}{n^2}\right).$$

Hence $\frac{1}{n}\operatorname{Tr}(\Gamma^h_{n,d})$ *converges to* $\beta_{h,d}$ *almost surely.*

Proof. The proof of this Lemma uses ideas from Bryc, Dembo and Jiang (2006). However one needs to argue slightly differently as the inputs of the matrix are no longer independent. We note that

$$\begin{aligned}
\mathbb{E}[\tfrac{1}{n}\operatorname{Tr}(\Gamma^h_{n,d}) - \tfrac{1}{n}\mathbb{E}[\operatorname{Tr}(\Gamma^h_{n,d})]^4 &= \tfrac{1}{n^{4h+4}}\mathbb{E}\Big[\sum_{\pi:\pi_h=\pi_0}\Big(\prod_{i=1}^h Y^{(n)}_{\pi_{i-1},\pi_i} - \mathbb{E}\Big[\prod_{i=1}^h Y^{(n)}_{\pi_{i-1},\pi_i}\Big]\Big)\Big]^4 \\
&= \tfrac{1}{n^{4h+4}}\mathbb{E}\Big[\sum_{(\mathbf{t},\pi)\in\mathcal{A}}\Big(X_{(\mathbf{a}(\mathbf{t},\pi))}\mathbb{I}_{(\mathbf{a}(\mathbf{t},\pi))} - \mathbb{E}\Big[X_{(\mathbf{a}(\mathbf{t},\pi))}\mathbb{I}_{(\mathbf{a}(\mathbf{t},\pi))}\Big]\Big)\Big]^4 \\
&= \tfrac{1}{n^{4h+4}}\sum_{\substack{\mathbf{t}^i,\pi^i\\ i=1,\dots,4}}\mathbb{E}\Big[\prod_{i=1}^4\Big(X_{(\mathbf{a}(\mathbf{t}^i,\pi^i))}\mathbb{I}_{(\mathbf{a}(\mathbf{t}^i,\pi^i))} - \mathbb{E}\Big[X_{(\mathbf{a}(\mathbf{t}^i,\pi^i))}\mathbb{I}_{(\mathbf{a}(\mathbf{t}^i,\pi^i))}\Big]\Big)\Big].
\end{aligned}$$

Suppose $\mathbf{m}^i = (m^i_1,\dots,m^i_{2h}) \in \{1,\dots,2n\}^{2h}; i=1,\dots,4$. We say $(\mathbf{m}^1,\dots,\mathbf{m}^4)$ are

- *jointly matched* if for any $m^i_j, (j\le 2h, i\le 4), \exists m^{i'}_{j'}, (j'\le 2h, i'\le 4)$ such that $|m^i_j - m^{i'}_{j'}| \le d$.
- *cross matched* if for any $i(\le 4), \exists j, j'(\le 2h)$ and $i'\neq i$ such that $|m^i_j - m^{i'}_{j'}| \le d$.

If $\mathbf{m}^i = (m^i_1,\dots,m^i_{2h})$ are not cross matched, say $\mathbf{m^1}\cap\mathbf{m^i} = \phi, \forall i\ge 2$, then

$$\begin{aligned}
\mathbb{E}[\textstyle\prod_{i=1}^4(X_{\mathbf{m}^i} - \mathbb{E}(X_{\mathbf{m}^i}))] &= \mathbb{E}[X_{\mathbf{m}^1} - \mathbb{E}(X_{\mathbf{m}^1})]\mathbb{E}[\textstyle\prod_{i=2}^4(X_{\mathbf{m}^i} - \mathbb{E}[X_{\mathbf{m}^i}])] \\
&= 0.
\end{aligned}$$

If $\mathbf{m}^i = (m^i_1,\dots,m^i_{2h})$ are not jointly matched, say m^1_1 appears only once then

$$\begin{aligned}
\mathbb{E}[\textstyle\prod_{i=1}^4(X_{\mathbf{m}^i} - \mathbb{E}(X_{\mathbf{m}^i}))] &= \mathbb{E}[X_{\mathbf{m^1}}\textstyle\prod_{i=2}^4(X_{\mathbf{m}^i} - \mathbb{E}(X_{\mathbf{m}^i}))] \\
&= \mathbb{E}[X_{m^1_1}]\mathbb{E}[\textstyle\prod_{j=2}^h X_{m^1_j}\prod_{i=2}^4(X_{\mathbf{m}^i} - \mathbb{E}(X_{\mathbf{m}^i}))] \\
&= 0.
\end{aligned}$$

Hence,

$$\mathbb{E}[\frac{1}{n}\operatorname{Tr}(\Gamma^h_{n,d}) - \frac{1}{n}\mathbb{E}(\operatorname{Tr}(\Gamma^h_{n,d}))]^4 = \frac{1}{n^{4h+4}}\sum_{\substack{(\mathbf{t}^i,\pi^i)\in\mathcal{A}\\ \mathbf{a}(\mathbf{t}^i,\pi^i)\text{ are jointly}\\ \text{and cross matched}\\ i=1,\dots,4}}\mathbb{E}[\prod_{i=1}^4(X_{(\mathbf{a}(\mathbf{t}^i,\pi^i)}\,\mathbb{I}_{(\mathbf{a}(\mathbf{t}^i,\pi^i)} - \mathbb{E}[X_{(\mathbf{a}(\mathbf{t}^i,\pi^i))}\mathbb{I}_{(\mathbf{a}(\mathbf{t}^i,\pi^i)}].$$

Since X_i's are uniformly bounded, it is enough to show, $\#\{(\mathbf{t}^1,\pi^1),\dots,(\mathbf{t}^4,\pi^4)\}$ such that $(\mathbf{a}(\mathbf{t}^1,\pi^1),\dots,\mathbf{a}(\mathbf{t}^4,\pi^4))$ are jointly matched and cross matched is $O(n^{4h+2})$.

Fix $\{\mathbf{a}(\mathbf{t}^i,\pi^i)|1\le i\le 4\}$ and consider the graph with $8h$ vertices $(u_1^1,...,u_{2h}^1,...,u_1^4,...,u_{2h}^4)$ with edges $u_l^i, u_{l'}^{i'}$ iff $|\mathbf{a}(\mathbf{t}^i,\pi^i)_l-\mathbf{a}(\mathbf{t}^{i'},\pi^{i'})_{l'}|\le d$.

Since $\{\mathbf{a}(\mathbf{t}^i,\pi^i)\}_{1\le i\le 4}$ are jointly matched, number of connected components in such a graph is at most $4h$.

Claim 3. $\#\{(\mathbf{t}^i,\pi^i)|1\le i\le 4\}$ which induces such graphs is $O(n^{4h+2})$.

Proof of Claim 3. We split the proof into three cases.

(a) First assume that the number of connected components is smaller or equal to $(4h-2)$. For each connected component, there are at most $O(n)$ many choices of selecting the corresponding elements of $\mathbf{a}(\mathbf{t}^i,\pi^i)$ and hence the number of ways to choose $\mathbf{a}(\mathbf{t}^i,\pi^i),(1\le i\le 4)$ is at most $O(n^{4h-2})$. Hence the claim is proved for this case.

(b) Next assume that there are $(4h-1)$ many components in the induced graph. Then there will be either one component with four vertices or there will be two components with three vertices each. Upon reordering if necessary, it can be easily checked that $\mathbf{a}^1=\mathbf{a}(\mathbf{t}^1,\pi^1)$ has an index a_{i*}^1 which does not match with any element in $\mathbf{a}^1$.

Now if $\{u_l^i,u_{l'}^{i'}\}$ is an edge, consider all possible relations of the form

$$\mathbf{a}(\mathbf{t}^i,\pi^i)_l-\mathbf{a}(\mathbf{t}^{i'},\pi^{i'})_{l'}=\pm k, |k|\le d$$

and all possibilities

$$b_j^i(\pi_{j-1}^i-\pi_j^i)=|\pi_{j-1}^i-\pi_j^i|.$$

We shall show that for every possible combination of relations $\#\{\mathbf{a}(\mathbf{t}^i,\pi^i)|i\le 4\}$ inducing the graph is $O(n^{4h-2})$ and hence (since there are finitely many such combinations)

$$\#\{(\mathbf{t}^i,\pi^i)|1\le i\le 4\}=O(n^{4h+2}).$$

By a slight abuse of notation let the *generating vertices* be a subset of $\{a_j^i\}$, chosen one from each connected component such that whenever a_j^i and $a_{j'}^{i'}$ belong to the same connected component, a_i^j belongs to the set of generating vertices iff either $i=i'$ and $j<j'$ or $i'>i$. Now let us choose the elements in $\mathbf{a}(\mathbf{t}^1,\pi^1)$ from left to right.

For $i<i^*$ if $a_i^1\in G$ then we can choose it in $O(n)$ ways and if $a_i^1\notin G$ then we can choose it in $O(1)$ ways. Next move on to $a_{i^*+1}^1$.

If it is a generating vertex we can choose it in $O(n)$ ways and otherwise in $O(1)$ ways. In this way complete all the choices for $\mathbf{a}(\mathbf{t}^1,\pi^1)$ except $a_{i^*}^1$.

Note that for a valid choice of $\mathbf{a}^1$ we must have $\sum_{j=1}^h b_j(a_{j+h}^1-a_j^1)=0$. This restriction automatically fixes $a_{i^*}^1$. Thus number of free choices reduces by one, and that implies

$$\#\{\mathbf{a}(\mathbf{t}^i,\pi^i)|i\le 4\}=O(n^{4h-1-1})=O(n^{4h-2}).$$

(c) Finally assume that there are $4h$ many connected components. Upon reordering it can be checked that there exists i^* and j^* such that $a_{i^*}^1$ does not match with any element in $\mathbf{a}^1$ and $a_{j^*}^2$ does not match with any element in $\mathbf{a}^1$ and $\mathbf{a}^2$. Now arguing as above it can be seen that number of free choices can be reduced by two and hence the claim is proved completely.

The Lemma follows immediately from this. □

3.3.6 Verification of Carleman's condition (C3) for Theorem 1 (a) (d finite)

Lemma 4. *The sequence* $\{\beta_{h,d}\}_{h\geq 0}$ *satisfies Carleman's condition and hence defines a unique probability distribution on* $\mathbb{R}$.

Proof. We recall the formulae for $p_{\mathbf{k}}^{(d)}$ and $\beta_{h,d}$ from (3.27) and (2.4). Now the number of ways of choosing the partition $\{1,...,2h\} = \cup_{l=1}^{h}\{i_l, j_l\}$ for $\mathbf{a}(\mathbf{t},\pi)$ is $\frac{(2h)!}{2^h h!}$. Hence

$$p_{\mathbf{k}}^{(d)} \leq \lim_n \frac{1}{n^{h+1}} \frac{(2h)!}{2^h h!} \times \frac{h!}{k_0! \ldots k_d!} \times 8^h n^{h+1} = \frac{4^h (2h)!}{h!} \frac{h!}{k_0! \cdots k_d!}.$$

Hence we have,

$$\begin{aligned} |\beta_{h,d}| &\leq \sum_{S_{h,d}} \frac{4^h(2h)!}{h!} \frac{h!}{k_0! \ldots k_d!} \prod_{i=0}^{d} |\gamma_{X^{(d)}}(i)|^{k_i} \\ &\leq \frac{4^h(2h)!}{h!} (\sum_{j=0}^{d} \sum_{k=0}^{d-j} |\theta_k \theta_{k+j}|)^h \leq \frac{4^h(2h)!}{h!} (\sum_{k=0}^{d} |\theta_k|)^{2h}. \end{aligned} \tag{3.30}$$

The above bound easily implies that $\sum_{h\geq 0} \beta_{2h,d}^{-\frac{1}{2h}} = \infty$ i.e. Carleman's condition is satisfied. This completes the proof of Theorem 1 (a). □

3.4 Proof of Theorem 1 (b) (infinite order case)

Fix $\varepsilon > 0$. Choose d such that $\sum_{k\geq d+1} |\theta_k| \leq \varepsilon$.

First we assume $\{\varepsilon_t\}$ *is i.i.d.* As earlier, define

$$A_n = \frac{1}{\sqrt{n}} \begin{bmatrix} 0 & X_1 & X_2 & \ldots & X_{n-1} & X_n & 0 & \ldots & 0 \\ 0 & 0 & X_1 & \ldots & X_{n-2} & X_{n-1} & X_n & \ldots & 0 \\ & & & & \vdots & & & & \\ 0 & 0 & 0 & \ldots & 0 & X_1 & X_2 & \ldots & X_n \end{bmatrix}_{n\times 2n}$$

so that

$$\begin{aligned} (A_n)_{i,j} &= X_{j-i}, \text{ if } 1 \leq j-i \leq n \\ &= 0, \text{ otherwise.} \end{aligned}$$

For convenience we will write

$$\Gamma_n(X) = \Gamma_n.$$

Clearly, $\Gamma_n = A_n A_n^T$ and we have

$$d_{BL}^2(F^{\Gamma_{n,d}}, F^{\Gamma_n}) \leq \frac{2}{n^2} \operatorname{Tr}\left[\Gamma_{n,d} + \Gamma_n\right] \operatorname{Tr}\left[(A_{n,d} - A_n)(A_{n,d} - A_n)^T\right].$$

Now, by *ergodic theorem*, almost surely,

$$\tfrac{1}{n}\left[\mathrm{Tr}(\Gamma_{n,d}+\Gamma_n)\right] \;= \frac{1}{n}\left[\sum_{t=1}^{n} X_{t,d}^2 + \sum_{t=1}^{n} X_t^2\right] \to \mathbb{E}[X_{t,d}^2 + X_t^2] \leq 2\sum_{k=0}^{\infty}\theta_k^2.$$

Similarly, almost surely,

$$\tfrac{1}{n}\,\mathrm{Tr}[(A_{n,d}-A_n)(A_{n,d}-A_n)^T] \;= \tfrac{1}{n}\sum_{t=1}^{n}(X_{t,d}-X_t)^2 \to \mathbb{E}[X_{t,d}-X_t]^2 \leq \sum_{k=d+1}^{\infty}\theta_k^2 \leq \varepsilon^2.$$

Hence almost surely

$$\limsup_n d_{BL}^2(F^{\Gamma_{n,d}}, F^{\Gamma_n}) \leq 2(\sum_{k=0}^{\infty}|\theta_k|)^2\varepsilon^2. \tag{3.31}$$

Now $F^{\Gamma_{n,d}} \xrightarrow{w} F_d$ almost surely. Since d_{BL} metrizes weak convergence of probability measures (on complete separable metric spaces, in particular on $\mathbb{R}$) we have as $n \to \infty$,

$$d_{BL}(F^{\Gamma_{n,d}}, F_d) \longrightarrow 0, \text{ almost surely.}$$

Using

$$d_{BL}(F^{\Gamma_n}, F^{\Gamma_m}) \leq d_{BL}(F^{\Gamma_n}, F^{\Gamma_{n,d}}) + d_{BL}(F^{\Gamma_{n,d}}, F^{\Gamma_{m,d}}) + d_{BL}(F^{\Gamma_{m,d}}, F^{\Gamma_m}),$$

the fact that $\{F^{\Gamma_{n,d}}\}_{n\geq 1}$ is Cauchy with respect to d_{BL} almost surely, and (3.31), we get

$$\limsup_{m,n} d_{BL}(F^{\Gamma_n}, F^{\Gamma_m}) \leq 2\sqrt{2}(\sum_{k=0}^{\infty}|\theta_k|)\varepsilon.$$

Hence $\{F^{\Gamma_n}\}_{n\geq 1}$ is Cauchy with respect to d_{BL} almost surely. Since d_{BL} is complete, there exists a probability measure F on $\mathbb{R}$ such that

$$F^{\Gamma_n} \xrightarrow{w} F \text{ almost surely.}$$

Further

$$d_{BL}(F_d, F) = \lim_n d_{BL}(F^{\Gamma_{n,d}}, F^{\Gamma_n}) \leq \sqrt{2}(\sum_{k=0}^{\infty}|\theta_k|)\varepsilon,$$

and hence

$$F_d \xrightarrow{w} F \text{ as } d \to \infty.$$

Since $\{F_d\}$ are nonrandom, we conclude that F is also nonrandom.

Now if $\{\varepsilon_t\}$ is not i.i.d. but independent and uniformly bounded by some $C > 0$ then the above proof is even simpler. One has to simply note that

$$\limsup_n \frac{1}{n}\sum_{t=1}^{n}\varepsilon_{t-k}^2 \leq C^2.$$

We omit the details. This completes the proof of the first part.

To show convergence of the moments $\{\beta_{h,d}\}$, we note that under the assumption of summability of the coefficients, (3.30) yields

$$\sup_d |\beta_{h,d}| \leq c_h := \frac{4^h(2h)!}{h!}(\sum_{k=0}^{\infty}|\theta_k|)^{2h} < \infty, \ \forall h \geq 0. \tag{3.32}$$

Thus we have uniform integrability of all powers of $A_d \sim F_d$. Since $F_d \stackrel{w}{\to} F$, we thus conclude

$$\beta_h = \int x^h \, dF = \lim_d \int x^h \, dF_d = \lim_{d\to\infty} \beta_{h,d}.$$

This completes the proof of (b). We note that $|\beta_h| \leq c_h$ from which we can show that $\{\beta_h\}_{h\geq 0}$ satisfies Carleman's condition, hence the moment sequence $\{\beta_h\}$ uniquely determines the distribution F. □

3.5 Proof of Theorem 1 (c)

3.5.1 Proof of moment ordering

We first prove the following.

Lemma 5. *For $d \geq 0$ and $k_0, \ldots, k_d \geq 0$,*

$$p^{(d)}_{k_0,\ldots,k_d} = p^{(d+1)}_{k_0,\ldots,k_d,0}.$$

Proof. Consider a graph G with $2h$ vertices with h connected components and two vertices in each component. Let

$$\mathcal{M} = \{\mathbf{a} : \ \mathbf{a} \text{ is minimal } d \text{ matched, induces } G \text{ and } |a_x - a_y| = d+1 \text{ for some } x, y \text{ belonging to distinct components of } G\}.$$

Then one can easily argue that $|\mathcal{M}| = O(n^{h-1})$ and consequently

$$\#\{(\mathbf{t},\pi) \in \mathcal{A} | \ \mathbf{a}(\mathbf{t},\pi) \in \mathcal{M}\} = O(n^h).$$

Hence

$$\begin{aligned} & p^{(d)}_{k_0,\ldots,k_d} \\ &= \lim_{n\to\infty} \frac{1}{n^{h+1}} \# \Big\{ (\mathbf{t},\pi) \in \mathcal{A} \ | \ \ \mathbf{a}(\mathbf{t},\pi) \text{ is minimal } d \text{ matched with partition } \{1,...,2h\} = \cup_{l=1}^h \{i_l, j_l\} \\ & \qquad \text{and there are exactly } k_s \text{many } l\text{'s for which} \\ & \qquad |\mathbf{a}(\mathbf{t},\pi)(i_l) - \mathbf{a}(\mathbf{t},\pi)(j_l)| = s, s = 0,...,d, \ \mathbb{I}_{\mathbf{a}(\mathbf{t},\pi)} = 1 \text{ and} \\ & \qquad |\mathbf{a}(\mathbf{t},\pi)(x) - \mathbf{a}(\mathbf{t},\pi)(y)| \geq d+2 \text{ if } x, y \text{ belong to} \\ & \qquad \text{different partition blocks} \Big\} \\ &= p^{(d+1)}_{k_0,\ldots,k_d,0}. \end{aligned}$$

□

When $\theta_0, \ldots, \theta_d \geq 0$ and $d \geq 1$

$$\begin{aligned}\beta_{h,d} &\geq \sum_{S_{h,d-1}} p^{(d)}_{k_0,\ldots,k_{d-1},0} \prod_{i=0}^{d-1} [\gamma_{X^{(d)}}(i)]^{k_i} \\ &\geq \sum_{S_{h,d-1}} p^{(d-1)}_{k_0,\ldots,k_{d-1}} \prod_{i=0}^{d-1} [\gamma_{X^{(d-1)}}(i)]^{k_i} \quad = \beta_{h,d-1}.\end{aligned}$$

□

Remark 5. *(Counter example for ordering of moments) If Assumption B(a) is violated, then the moment ordering need not hold. Consider an $MA(2)$ process with parameters $\theta_0,\ \theta_1,\ \theta_2$ and an $MA(1)$ with parameter set $\theta_0,\ \theta_1$. Using Lemma 9 we note that*

$$\beta_{2,2} = p^{(2)}_{2,0,0}(\theta_0^2 + \theta_1^2 + \theta_2^2)^2 + p^{(2)}_{0,2,0}(\theta_0\theta_1 + \theta_1\theta_2)^2 + p^{(2)}_{0,0,2}\theta_0^2\theta_2^2 \tag{3.33}$$

and

$$\beta_{2,1} = p^{(1)}_{2,0}(\theta_0^2 + \theta_1^2)^2 + p^{(1)}_{0,2}\theta_0^2\theta_1^2. \tag{3.34}$$

Using Lemma 5 we get $p^{(2)}_{0,2,0} = p^{(1)}_{0,2}$. Further, it is also not hard to verify that $p^{(2)}_{0,0,2} = p^{(2)}_{0,2,0}$. Thus $\beta_{2,2} \geq \beta_{2,1}$ iff

$$p^{(1)}_{2,0}[\theta_2^4 + 2(\theta_0^2 + \theta_1^2)\theta_0^2] + p^{(1)}_{0,2}[\theta_1^2\theta_2^2 + 2\theta_0\theta_1^2\theta_2 + \theta_0^2\theta_2^2] \geq 0 \tag{3.35}$$

Now taking $\theta_2 = -\kappa\theta_0$ where $\kappa > 0$ and $\theta_0, \theta_1 > 0$ after some simplification we get

$$\beta_{2,2} \geq \beta_{2,1} \Leftrightarrow \kappa(p^{(1)}_{2,0}(\kappa^2 + 2) + p^{(1)}_{0,2})\theta_0^2 + (2p^{(1)}_{2,0}\kappa + p^{(1)}_{0,2}(\kappa - 2))\theta_1^2 \geq 0 \tag{3.36}$$

After solving a linear equation for κ, it is easily seen that there exists a $\kappa^* > 0$ such that if $\kappa \in (0, \kappa^*)$, then coefficient of θ_1 will be negative. Hence fixing some arbitrary value of $\theta_0 > 0$, and $\kappa \in (0, \kappa^*)$ one can increase the value of θ_1 arbitrarily to get $\beta_{2,2} < \beta_{2,1}$.

3.5.2 Proof of unbounded support of F_d

We use the approach of Bryc, Dembo and Jiang (2006). Let

$\mathcal{W} = \{w = w_1w_2 :\ |w_1| = 2h = |w_2|;\ w, w_1, w_2$ are zero pair matched; $w_1[x]$ matches with $w_1[y]$ iff $w_2[x]$ matches with $w_2[y]\}$.

Then

$$\beta_{2h,d} \geq [\gamma_{X^{(d)}}(0)]^{2h} p_{2h,0,\ldots,0} \geq [\gamma_{X^{(d)}}(0)]^{2h} \sum_{w \in \mathcal{W}} \lim_n n^{-(2h+1)} |\Pi^*(w)|. \tag{3.37}$$

For $w = w_1w_2 \in \mathcal{W}$, let $\{1, \ldots, 2h\} = \cup_{i=1}^h (i_s, j_s)$ be the partition corresponding to w_1. Then

$$\begin{aligned}\lim_n \tfrac{|\Pi^*(w)|}{n^{2h+1}} \geq \lim_n \tfrac{1}{n^{2h+1}} \#\{(\mathbf{t}, \pi) :\ & t_{i_s} = t_{j_s} \text{ and } \pi_{i_s} - \pi_{i_s-1} = \pi_{j_s-1} - \pi_{j_s} \text{ for } 1 \leq s \leq h; \\ & t_j + |\pi_j - \pi_{j-1}| \leq n, \text{ for } 1 \leq j \leq 2h\}.\end{aligned} \tag{3.38}$$

Using the relations

$$\pi_{i_s} - \pi_{i_s-1} = \pi_{j_s-1} - \pi_{j_s},\ 1 \le s \le h,$$

we can construct the generating vertices set

$$\{v_0, \dots, v_h\} \subset \{\pi_0, \dots, \pi_{2h}\}$$

and write

$$\pi_i = \lambda_{\mathbf{i}}(\mathbf{v}),\ (\mathbf{v} = \{v_0, \dots, v_h\}) \text{ for } i = 0, \dots, 2h.$$

From the relations arising out of the word w_1, construct a generating set $\{u_1, \dots, u_h\} \subset \{t_1, \dots, t_{2h}\}$ to write $t_i = \beta_i(\mathbf{u})$ where $\mathbf{u} = (u_1, \dots, u_h)$, (in fact $\beta_{i_s}(\mathbf{u}) = \beta_{j_s}(\mathbf{u}) = u_s,\ 1 \le s \le h$).

It is easy to see that

$$\lambda_{\mathbf{i}}(\mathbf{v}) = \sum_{a=0}^{h} \lambda_{\mathbf{ia}} v_a \text{ where } \lambda_{\mathbf{ia}} \in \{-1, 0, 1\} \text{ and } \sum_{a=0}^{h} \lambda_{\mathbf{ia}} = 1\ \forall i.$$

From (3.38) we get for i.i.d Uniform random variables $U_1, \dots, U_h, V_0, \dots, V_h$,

$$\begin{aligned}
\lim_n \frac{|\Pi_*(w)|}{n^{2h+1}} &\ge \mathbb{P}\left(\lambda_{\mathbf{i}}(\mathbf{V}) \in (0,1), \beta_i(\mathbf{U}) + |\lambda_{\mathbf{i}}(\mathbf{V}) - \lambda_{\mathbf{i-1}}(\mathbf{V})| \le 1,\ i = 1, \dots, 2h\right) \\
&\ge \mathbb{P}\left(\lambda_{\mathbf{i}}(\mathbf{V}) \in (\frac{1}{4}, \frac{3}{4}), \beta_i(\mathbf{U}) \in (0, \frac{1}{2}),\ i = 1, \dots, 2h\right) \\
&= 2^{-h}\mathbb{P}\left(\lambda_{\mathbf{i}}(\mathbf{V}) \in (\frac{1}{4}, \frac{3}{4}),\ i = 1, \dots, 2h\right) \\
&= 2^{-h}\mathbb{P}\left(\sum_{a=0}^{h} \lambda_{\mathbf{ia}} Y_a \in (-\frac{1}{4}, \frac{1}{4}),\ i = 1, \dots, 2h\right), \quad [Y_a = V_a - \frac{1}{2}, i = 0, \dots, h] \\
&\ge 2^{-h}\mathbb{P}\left(\sum_{a=0}^{h} \lambda_{\mathbf{ia}} Y_a \in (-\frac{1}{4}, \frac{1}{4}),\ i = 1, \dots, 2h \Big| A\right) \mathbb{P}(A).
\end{aligned}$$

Here $A = \bigcap_{j=0}^{h} \{|Y_j| \le \frac{1}{2\varepsilon(h+1)}\}$, $\varepsilon > 0$ being fixed and h is large so that $\varepsilon(h+1) > 1$. Conditional on A, Y_j are i.i.d. $Uniform[-\frac{1}{2\varepsilon(h+1)}, \frac{1}{2\varepsilon(h+1)}]$. Then

$$\begin{aligned}
\mathbb{P}\left(\sum_{a=0}^{h} \lambda_{\mathbf{ia}} Y_a \notin (-\frac{1}{4}, \frac{1}{4}),\ \text{for some } 1 \le i \le 2h \Big| A\right) &\le \sum_{i=1}^{2h} \mathbb{P}\left(\sum_{a=0}^{h} \lambda_{\mathbf{ia}} Y_a \notin (-\frac{1}{4}, \frac{1}{4}) \Big| A\right) \\
&\le \sum_{i=1}^{2h} 2\exp(-\frac{(\varepsilon(h+1))^2}{8k_i'}) \\
&\qquad (\text{ where } k_i' = \#\{a |\ \lambda_{\mathbf{ia}} \ne 0\}) \\
&\le 4h\exp(-\frac{\varepsilon^2(h+1)}{8})
\end{aligned}$$

where the second inequality follows from Hoeffding's inequality.

Since the last expression tends to zero, we have for large enough h,

$$P\left(\sum_{a=0}^{h}\lambda_{\mathbf{ia}}Y_a \in (-\frac{1}{4},\frac{1}{4}),\ i=0,\ldots,2h\Big|A\right) \geq \frac{1}{2}.$$

Since $|\mathcal{W}| = \frac{2h!}{2^h h!}$ from (3.37), we have for large enough h and constants C, C', C''

$$\beta_{2h,d} \geq C^{2h}\frac{2h!}{2^h h!}\mathbb{P}(A) \geq C'^{2h}\frac{h^h}{(\varepsilon(h+1))^{h+1}} \text{ (by Stirling's approximation)}$$

and hence

$$\limsup_h \beta_{2h,d}^{\frac{1}{2h}} \geq C''\varepsilon^{-\frac{1}{2}}.$$

Letting ε go to zero we get unboundedness of the support of F_d.

Since $\{\beta_{h,d}\}$ increases to β_h, $\limsup_h \beta_h^{\frac{1}{h}} = \infty$, and hence F also has unbounded support.

3.6 Outline of the proof of Theorem 3

Here is an outline of the changes required to prove Theorem 1 for other values of α.

3.6.1 Proof of Theorem 3 for the case $0 < \alpha < 1$

Let $\beta_h(\Gamma_{n,d}^{\alpha,I})$ and $\beta_h(\Gamma_{n,d}^{\alpha,II})$ be the h^{th} moments respectively of the ESD of Type I and Type II band autocovariance matrices with parameter α. We begin by noting that the expression for these contain an extra indicator term $\prod_{i=1}^{h}\mathbb{I}(|\pi_{i-1}-\pi_i| \leq m_n)$ and $\prod_{i=1}^{h}\mathbb{I}(1 \leq \pi_i \leq m_n)$ respectively. For Type II band autocovariance matrices since there are m_n eigenvalues instead of n, the normalising denominator is now m_n. Hence

$$\beta_h(\Gamma_{n,d}^{\alpha,I}) = \frac{1}{n^{h+1}}\sum_{\substack{1\leq \pi_0,\ldots,\pi_h\leq n\\ \pi_h=\pi_0}}\Big[\prod_{j=1}^{h}\Big(\sum_{t_j=1}^{n}X_{t_j,d}X_{t_j+|\pi_j-\pi_{j-1}|,d}\mathbb{I}_{(t_j+|\pi_j-\pi_{j-1}|\leq n)}\Big)\Big] \times \prod_{i=1}^{h}\mathbb{I}(|\pi_{i-1}-\pi_i| \leq m_n) \tag{3.39}$$

and

$$\frac{m_n}{n}\beta_h(\Gamma_{n,d}^{\alpha,II}) = \frac{1}{n^{h+1}}\sum_{\substack{1\leq \pi_0,\ldots,\pi_h\leq n\\ \pi_h=\pi_0}}\Big[\prod_{j=1}^{h}\Big(\sum_{t_j=1}^{n}X_{t_j,d}X_{t_j+|\pi_j-\pi_{j-1}|,d}\mathbb{I}_{(t_j+|\pi_j-\pi_{j-1}|\leq n)}\Big)\Big] \times \prod_{i=1}^{h}\mathbb{I}(1 \leq \pi_i \leq m_n) \tag{3.40}$$

It is thus enough to establish the limits on the right side of the above expressions. and we can follow similar steps as in the proof of Theorem 1.
Since there are only some extra indicator terms, the negligibility of higher order edges and verification of (C2) and (C3) needs no new arguments. By the same logic, verification of (C1) is also similar except that there will be an extra term coming in the expression for $p_w^{(d)}$ now. From (3.39) and (3.40) we note that the extra term $\prod_{j=2}^{h+1} \mathbb{I}(|\lambda^{\mathbf{b}}_{\mathbf{j+h-1}} - \lambda^{\mathbf{b}}_{\mathbf{j+h}}| \leq \alpha)$ will be inside the expectation for $\Gamma_{n,d}^{\alpha,I}$ and for $\Gamma_{n,d}^{\alpha,II}$ the corresponding term will be $\prod_{j=1}^{h+1} \mathbb{I}(\lambda^{\mathbf{b}}_{\mathbf{j+h}} \in (0,\alpha))$.
This would complete the proof for finite d.
To establish the result for $d = \infty$, note that the Type II band autocovariance matrices are $m_n \times m_n$ principal subminor of the original sample autocovariance matrices they are automatically nonnegative definite and we can write

$$\Gamma_n^{\alpha,II}(X^{(d)}) = (A_{n,d}^{\alpha,II})(A_{n,d}^{\alpha,II})^T$$

where $A_{n,d}^{\alpha,II}$ is the first m_n rows of $A_{n,d}$. Thus we can establish the connection between the limiting distribution between d finite and $d = \infty$ imitating the ideas in Theorem 1. However to prove the same conclusion for Type I band autocovariance matrices we can not apply Theorem 1 as these matrices may not be nonnegative definite. Thus here we proceed similarly as in Theorem 2.

Proof of unbounded support needs only minor changes. We omit the details.

When the input sequence satisfy only Assumption $A(a)$ noting that $\Gamma_{n,d}^{\alpha,II}$ are nonnegative definite and using the same technique as in Theorem 1 we can without of loss of generality assume Assumption $A(b)$ holds. This completes proof of this part. □

3.6.2 Proof of Theorem 3 for $\alpha = 0$, Type I band autocovariance matrix

Existence: When $\alpha = 0$, using (3.39) we note that we got the following expression instead of (3.24) which we got for $\alpha = 1$. Below $p_w^{(d),0,I}$ denote the limiting contribution of the word w for Type I band autocovariance matrix with band parameter $\alpha = 0$.

$$\begin{aligned} p_w^{(d),0,I} := \lim_n \frac{1}{n^{h+1}} |\Pi^{0,*}(w)| \quad = \quad & \lim_n \frac{1}{n^{h+1}} \sum_{\mathbf{b} \in \mathcal{B}(w)} \sum_{U^n \in \mathcal{N}_n^{h+1}} \mathbb{I}({\lambda^{\mathbf{b}}}_{2h+1}(U^n) + m_{2h+1} = {\lambda^{\mathbf{b}}}_{h+1}(U^n) + m_{h+1}) \\ & \times \prod_{j=1}^{2h} \mathbb{I}(\lambda^{\mathbf{b}}_{\mathbf{j}}(U^n) + m_j \in \mathcal{N}_n) \\ & \times \prod_{\substack{j=1 \\ j \notin \mathcal{T}}}^{h} \mathbb{I}(0 \leq b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(U^n) - \lambda^{\mathbf{b}}_{\mathbf{j+h}}(U^n) + m_{j+h-1} - m_{j+h}) \leq m_n) \\ & \times \prod_{j=1}^{h} \mathbb{I}(\lambda^{\mathbf{b}}_{\mathbf{j}}(U^n) + b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(U^n) - {\lambda^{\mathbf{b}}}_{j+h}(U^n) + m_{j+h-1} - m_{j+h}) \leq n) \end{aligned}$$

$$\times \prod_{j \in \mathcal{T}} \mathbb{I}(-m_n \leq n_j \leq 0). \tag{3.41}$$

Note that if for a word w, $\lambda^{\mathbf{b}}_{\mathbf{j+h-1}} \neq \lambda^{\mathbf{b}}_{\mathbf{j+h}}$ for some j then the third indicator term in (3.41) will go to zero as $n \to \infty$ and thus limiting contribution from that word will be 0. Thus only those words w for which $\lambda^{\mathbf{b}}_{\mathbf{h+1}} = \lambda^{\mathbf{b}}_{\mathbf{j+h}}$ for all $j \in \{1, 2, \ldots, h+1\}$ may contribute nonzero quantity in the limit. This condition also implies that, for such words no π_i belongs to the generating set except π_0. This observation together with Lemma 6 and the expression for limiting moments for $\Gamma_n(X)$ shows that $w \in \mathcal{W}^h_0$ may contribute nonzero quantity, where

$$\mathcal{W}^h_0 = \{w : \ |w| = 2h, \ w[i] \text{ matches with } w[i+h], \ n_i \leq 0, \ i = 1, 2, \ldots, h\}.$$

Now note that if $w \in \mathcal{W}^h_0$ then $\mathcal{T} = \{h+1, h+2, \ldots, 2h\}$ and thus third indicator term in (3.41) vanishes. So we get a modified expression for the limiting contribution.

$$\begin{aligned} p_w^{(d),0,I} &= \lim_n \frac{1}{n^{h+1}} \sum_{\mathbf{b} \in \mathcal{B}(w)} \sum_{U^n \in \mathcal{N}_n^{h+1}} \mathbb{I}(m_{2h+1} = m_{h+1}) \times \prod_{j=1}^{h} \mathbb{I}(-m_n \leq n_j \leq 0) \\ &\times \prod_{j=1}^{2h} \mathbb{I}(\lambda^{\mathbf{b}}_{\mathbf{j}}(U^n) + m_j \in \mathcal{N}_n) \times \prod_{j=1}^{h} \mathbb{I}(\lambda^{\mathbf{b}}_{\mathbf{j}}(U^n) - n_j \leq n). \end{aligned} \tag{3.42}$$

For $d = 0$ note that $|\mathcal{W}^h_0| = 1$ for every h one can easily check that the contribution from that word is 1. Thus $\beta^0_{h,0} = \theta_0^{2h}$ and as a consequence, the LSD is $\delta_{\theta_0^2}$.

Now let us consider any $0 < d < \infty$. Note that for any d finite the last two indicators in the above expression go to 1 as $n \to \infty$. The second indicator becomes $\prod_{j=1}^{h} \mathbb{I}(n_j \leq 0)$ if $m_n \geq d$ as $|n_j| \leq d$. Thus for any $w \in \mathcal{W}^h_0$ the limiting contribution $p_w^{(d),0,I}$ will be the number of $\mathbf{b} \in \mathcal{B}(w)$ such that $\sum n_i b_i = 0$. Thus we obtain an expression for the limit moments.

Now assume $d = \infty$. To prove the existence of LSD we proceed as earlier. We already noted that Type I band autocovariance matrices need not be non negative definite. Thus to prove the connection of limits for d finite and $d = \infty$ we again use the technique from Theorem 2. Hence we have proved that the LSD exists.

Identification of the LSD: Now it remains to argue that the limit we obtained is same as $f_X(U)$. For $d = 0$ we have already identified the limit for Type I autocovariance matrices and it is trivial to check it is same as $f_X(U)$.

Now consider any $0 < d < \infty$. Note that in the proof above we did not use the fact that $m_n \to \infty$ and we further note that for any sequence $\{m_n\}$ the limit we obtained above will be same whenever $\liminf_{n \to \infty} m_n \geq d$. So in particular the limit will be same if we choose another sequence $\{m'_n\}$ such that $m'_n = d$ for all n. Now we argue that limiting spectral distribution of these sequence matrices is same as with that of Σ_n. To show this we make another use of Bounded Lipschitz metric. Let $\Gamma^I_{n',d}$ denote the Type I band autocovariance matrix where we put 0 instead of $\hat{\gamma}_{X^{(d)}}(k)$ whenever $k > m'_n$ and let $\Sigma_{n,d}$ be the $n \times n$ matrix whose $(i,j)^{th}$ entry is the population autocovariance $\gamma_{X^{(d)}}(|i-j|)$. Now from Lemma 1 (a) we get

$$d_{BL}^2(F^{\Gamma^I_{n',d}}, F^{\Sigma_{n,d}}) \leq \frac{1}{n}\operatorname{Tr}(\Gamma^I_{n',d} - \Sigma_{n,d})^2 \;\leq\; 2(\hat{\gamma}_{X^{(d)}}(0) - \gamma_{X^{(d)}}(0))^2 + \cdots + 2(\hat{\gamma}_{X^{(d)}}(d) - \gamma_{X^{(d)}}(d))^2$$

For any j as $n \to \infty$, $\hat{\gamma}_{X^{(d)}}(j) \to \gamma_{X^{(d)}}(j)$ almost surely. Using the fact that d is finite, the right hand side of the above expression goes to 0 almost surely. This proves the claim for d finite.

First note that we already have

$$LSD(\Gamma^{0,I}_{n,d}) = LSD(\Sigma_{n,d}) := G_d \text{ and } LSD(\Gamma^{0,I}_{n,d}) \xrightarrow{w} LSD(\Gamma^{0,I}_{n}) \text{ as } d \to \infty.$$

Now it remains to proof for the case $d = \infty$. Thus enough to prove that

$$G_d \xrightarrow{w} G(= LSD(\Sigma_n)) \text{ as } d \to \infty$$

where Σ_n is the $n \times n$ matrix whose $(i,j)^{th}$ entry is $\gamma_X(|i-j|)$. To prove the above we make another use of Bounded Lipschitz metric. First let us define another sequence of $n \times n$ matrices $\bar{\Sigma}_{n,d}$ whose $(i,j)^{th}$ entry is $\gamma_X(|i-j|)$ if $|i-j| \leq d$ and otherwise 0. Note

$$d_{BL}^2(F^{\Sigma_{n,d}}, F^{\Sigma_n}) \leq 2d_{BL}^2(F^{\Sigma_{n,d}}, F^{\bar{\Sigma}_{n,d}}) + 2d_{BL}^2(F^{\bar{\Sigma}_{n,d}}, F^{\Sigma_n}) \tag{3.43}$$

Fix any $\varepsilon > 0$. Fix d_0 such that $\Big(\sum_{j=0}^{\infty}|\theta_j|\Big)^2\Big(\sum_{l=d+1}^{\infty}|\theta_l|\Big)^2 \leq \frac{\varepsilon^2}{32}$ for all $d \geq d_0$. Now using Lemma 1 (a),

$$\begin{aligned}
\limsup_n d_{BL}^2(F^{\Sigma_{n,d}}, F^{\bar{\Sigma}_{n,d}}) &\leq \limsup_n \frac{1}{n}\operatorname{Tr}(\Sigma_{n,d} - \bar{\Sigma}_{n,d})^2 \\
&\leq 2\Big[(\gamma_{X^{(d)}}(0) - \gamma_X(0))^2 + \cdots + (\gamma_{X^{(d)}}(d) - \gamma_X(d))^2\Big] \\
&= 2\sum_{j=0}^{d}\Big(\sum_{k=d-j+1}^{\infty}\theta_k\theta_{j+k}\Big)^2 \\
&\leq 2\Big(\sum_{l=d+1}^{\infty}|\theta_l|\sum_{j=0}^{\infty}|\theta_j|\Big)^2 \\
&= 2\Big(\sum_{j=0}^{\infty}|\theta_j|\Big)^2\Big(\sum_{l=d+1}^{\infty}|\theta_l|\Big)^2 \leq \frac{\varepsilon^2}{16}
\end{aligned} \tag{3.44}$$

and

$$d_{BL}^2(F^{\bar{\Sigma}_{n,d}}, F^{\Sigma_n}) \leq \limsup_n \frac{1}{n}\operatorname{Tr}(\bar{\Sigma}_{n,d} - \Sigma_n)^2 \tag{3.45}$$

$$\leq 2\sum_{j=d+1}^{\infty}\gamma_X(j)^2 = 2\Big(\sum_{k=0}^{\infty}|\theta_k|\Big)^2\Big(\sum_{j=d+1}^{\infty}|\theta_j|\Big)^2 \leq \frac{\varepsilon^2}{16} \tag{3.46}$$

Thus from (3.43), (3.44) and (3.46) for any $d \geq d_0$, we have

$$\limsup_n d_{BL}(F^{\Sigma_{n,d}}, F^{\Sigma_n}) \leq \varepsilon/2. \tag{3.47}$$

Since

$$d_{BL}(F^{G_d}, F^G) \leq d_{BL}(F^{G_d}, F^{\Sigma_{n,d}}) + d_{BL}(F^{\Sigma_{n,d}}, F^{\Sigma_n}) + d_{BL}(F^{\Sigma_n}, F^G)$$

we have for any $d \geq d_0$,

$$d_{BL}(F^{G_d}, F^G) \leq \varepsilon.$$

This completes the proof. $\square$

3.6.3 Proof of Theorem 3 for $\alpha = 0$, Type II band autocovariance matrix

Part of this proof will be different from the corresponding proof for Type I band autocovariance matrices, as the expressions for the h^{th} moment of the ESD of these matrices differ by a factor α_n and here $\alpha_n \to 0$.
First note that by Lemma 2 we need to consider only minimal matched terms. Let

$$G_t = \{t_i : \ t_i \in G\} \text{ and } G_\pi = \{\pi_i : \ \pi_i \in G\}.$$

Since $1 \leq \pi_i \leq m_n$ for all i, by similar arguments as in Lemma 2 we get

$$\text{number of choices of } \ \mathbf{a}(\mathbf{t}, \pi) = O(n^{|G_t|} m_n^{|G_\pi|}).$$

Thus for any word w such that $|G_t| < h$ the limiting contribution will be 0. Hence only words to contribute in this case will be those for which $|S_3(w)| = |S_4(w)| = h$ and from Lemma 6 we get that only words to contribute here are those belonging to $\mathcal{W}_0^h$. Therefore using same arguments as in the proof of Theorem 3, for Type I band autocovariance matrices for $\alpha = 0$ we obtain the same limit as we obtained there. All the remaining conclusions here follow from the proof for Type I band autocovariance matrices with parameter $\alpha = 0$. But here note that in order to achieve the same limit as obtained for Type I autocovariance matrices we need $m_n \to \infty$ even if d is finite.
To prove connection between the limiting distribution for d finite and $d = \infty$ we use same arguments as in Theorem 1 as these matrices are nonnegative definite.

$\square$

3.7 Proof of Theorem 4

We provide only an outline of the arguments. Parts of the proof will be borrowed from the proof of Theorem 3 for the case $\alpha = 0$. First fix any d fnite.
First note that since K is bounded, negligibility of higher order edges, verification of $(C2)$ and $(C3)$ is same as before. Following same technique as before we can also verify $(C1)$ and we note that for the tapered case the expression will be almost same as (3.41) but it will contain an extra factor here. Below $p_w^{(d),K}$ denote the limiting contribution from a word w.

$$\begin{aligned} p_w^{(d),K} := \lim_n \frac{1}{n^{h+1}} |\Pi^{K,*}(w)| \ &= \ \lim_n \frac{1}{n^{h+1}} \sum_{\mathbf{b} \in \mathcal{B}(w)} \sum_{U^n \in \mathcal{N}_n^{h+1}} \mathbb{I}(\lambda^{\mathbf{b}}{}_{2h+1}(U^n) + m_{2h+1} = \lambda^{\mathbf{b}}{}_{h+1}(U^n) + m_{h+1}) \\ &\times \prod_{j=1}^{2h} \mathbb{I}(\lambda_{\mathbf{j}}^{\mathbf{b}}(U^n) + m_j \in \mathcal{N}_n) \times \prod_{j \in \mathcal{T}} \mathbb{I}(-m_n \leq n_j \leq 0) \end{aligned}$$

$$\times \prod_{\substack{j=1 \\ j \notin \mathcal{T}}}^{h} \mathbb{I}(0 \le b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(U^n) - \lambda^{\mathbf{b}}_{\mathbf{j+h}}(U^n) + m_{j+h-1} - m_{j+h}) \le m_n)$$

$$\times \prod_{j=1}^{h} \mathbb{I}(\lambda^{\mathbf{b}}_{\mathbf{j}}(U^n) + b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(U^n) - \lambda^{\mathbf{b}}{}_{j+h}(U^n) + m_{j+h-1} - m_{j+h}) \le n)$$

$$\times \prod_{j=1}^{h} K\left(\frac{b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(U^n) - \lambda^{\mathbf{b}}{}_{j+h}(U^n) + m_{j+h-1} - m_{j+h})}{m_n}\right). \tag{3.48}$$

Following the arguments in Section 3.6.2 and simplifying (3.48) as before we get

$$p_w^{(d),K} = \lim_n \frac{1}{n^{h+1}} \sum_{\mathbf{b} \in \mathcal{B}(w)} \sum_{U^n \in \mathcal{N}_n^{h+1}} \mathbb{I}(m_{2h+1} = m_{h+1}) \times \prod_{j=1}^{h} \mathbb{I}(-m_n \le n_j \le 0)$$
$$\times \prod_{j=1}^{2h} \mathbb{I}(\lambda^{\mathbf{b}}_{\mathbf{j}}(U^n) + m_j \in \mathcal{N}_n) \times \prod_{j=1}^{h} \mathbb{I}(\lambda^{\mathbf{b}}_{\mathbf{j}}(U^n) - n_j \le n) \times \prod_{j=1}^{h} K\Big(\frac{-n_j}{m_n}\Big). \tag{3.49}$$

Since $m_n \to \infty$, $K(\cdot)$ is continuous at 0, $K(0) = 1$ and arguing as in Section 3.6.2, we get $p_w^{(d),0,I} = p_w^{(d),K}$ for every word w and thus the limiting distributions are same in both the cases.

For the case $d = \infty$ the arguments will be exactly similar as in Section 3.6.2. So we omit the details. This completes the proof. □

Remark 6. *Basak and Bose (2010) deal with two matrices which are actually weighted versions of the standard Toeplitz and Hankel matrices. The corresponding tapering function K there was unbounded which necessitated some special arguments.*

With the arguments of the above paper in mind, when K is unbounded, first one would argue that the d-pair matched words are the only words which continue to contribute to the limit. Then one needs appropriate generalizations of Lemma 5 and Lemma 6 of Basak and Bose (2010). Then using a generalized Holder inequality, one can tackle the limit moments. Finally using an approach similar to that in Section 2.6-2.8 of Basak and Bose (2010) one may conclude the existence of the LSD. Using similar techniques one may also conclude the limiting distribution will be $f_X(U)$ when $m_n \to \infty$ but $m_n/n \to 0$. However the extent of difficulty of carrying out these steps will depend on the conditions imposed on the function K.

3.8 Proof of Theorem 2

We give a brief description and omit the details. Define

$$Y_{i,j}^{(n)} = \sum_{t=1}^{n} X_t X_{t+|i-j|}.$$

In this case, $\lim_n \frac{1}{n^{h+1}} |\Pi^*(w)|$ is given by

$$\sum_{b \in \mathcal{B}(w)} \mathbb{P}\Big(\lambda^{\mathbf{b}}_{\mathbf{j}}(V) \in (0,1),\ 1 \le j \le 2h;\ b_j(\lambda^{\mathbf{b}}_{\mathbf{j+h-1}}(V) - \lambda^{\mathbf{b}}_{\mathbf{j+h+1}}(V)) \ge 0,\ 1 \le j \le h,\ j \notin \mathcal{J},$$

$$\lambda^{\mathbf{b}}_{\mathbf{2h+1}}(V) = \lambda^{\mathbf{b}}_{\mathbf{h+1}}(V), \Big) \times I(m_{h+1} = m_{2h+1}) \times I(n_j \le 0, j \in \mathcal{J}).$$

Comparing this with the corresponding expression for the sequence $\Gamma_{n,d}$, it follows that

$$\beta_{h,d} \le \beta^*_{h,d} \ \text{ if } \ \theta_j \ge 0, \ 0 \le j \le d.$$

Relation (3.30) holds with $\beta_{h,d}$ replaced by $\beta^*_{h,d}$. We can use this to prove tightness of $\{F^*_d\}$ under Assumption B(a) and then proceed to establish Carleman's condition. We omit the tedious details.

Since Γ^*_n and $\Gamma^*_{n,d}$ are no longer positive definite matrices we cannot imitate the idea used in the proof of Theorem 1(b). We proceed as follows instead. Note that

$$\mathbb{E}[\beta_h(\Gamma^*_n)] = \frac{1}{n^{h+1}} \mathbb{E}\Big[\sum_{(\mathbf{t},\pi)\in\mathcal{A}} \prod_{j=1}^{h} X_{t_j} \prod_{j=1}^{h} X_{t_j+|\pi_{j-1}-\pi_j|} \Big].$$

Write

$$X_{t_j} = \sum_{k_j\ge 0} \theta_{k_j} \varepsilon_{t_j-k_j} \ \text{ and } \ X_{t_j+|\pi_{j-1}-\pi_j|} = \sum_{k'_j\ge 0} \theta_{k'_j} \varepsilon_{t_j+|\pi_{j-1}-\pi_j|-k'_j}.$$

Then using the absolute summability Assumption B(b) and applying DCT we get

$$\mathbb{E}[\beta_h(\Gamma^*_n)] = \sum_{\substack{k_j,k'_j\ge 0\\ j=1,\ldots,h}} \prod_{j=1}^{h} (\theta_{k_j}\theta_{k'_j}) \frac{1}{n^{h+1}} \mathbb{E}\Big[\sum_{(\mathbf{t},\pi)\in\mathcal{A}} \prod_{j=1}^{h} \varepsilon_{t_j-k_j} \varepsilon_{t_j+|\pi_j-\pi_{j-1}|-k'_j} \Big].$$

Then using the fact that $\{\varepsilon_t\}_{t=1}^{\infty}$ are uniformly bounded and absolute summability of $\{\theta_k\}_{k=1}^{\infty}$ we note that it is enough to show that the limit below exists.

$$\lim_n n^{-(h+1)} \mathbb{E}\Big[\sum_{(\mathbf{t},\pi)\in\mathcal{A}} \prod_{j=1}^{h} (\varepsilon_{t_j-k_j} \varepsilon_{t_j+|\pi_j-\pi_{j-1}|-k'_j}) \Big].$$

One can proceed as in the proof of Theorem 1 to show that only pair matched words contribute and hence enough to argue that

$\lim n^{-(h+1)} \#\{(\mathbf{t},\pi) \in \mathcal{A} \ : \ \{t_j - k_j, \ t_j + |\pi_j - \pi_{j-1}| - k'_j, \ j = 1, \ldots, h\}$ is pair matched$\}$ exists. Now we again proceed as in the proof of Theorem 1 to show that the above limit exists. Of course there is some compatibility needed among $\{k_j, k'_j, \ j = 1, \ldots, h\}$, the word w and the signs b_i $(= \pm 1)$ to ensure that the condition $\pi_0 = \pi_h$ is satisfied. So the above limit will actually depend on $\{k_j, k'_j, \ j = 1, \ldots, h\}$.

We also note that

$$\lim_n \frac{1}{n^{h+1}} \sum_{\substack{w \text{ pair matched},\\ |w|=2h}} \#\{(\mathbf{t},\pi) \in \mathcal{A} \ : \ (t_j - k_j, \ t_j + |\pi_j - \pi_{j-1}| - k'_j)_{j=1,\ldots,h} \in \Pi(w)\} \le \frac{4^h (2h)!}{h!}.$$

From this it follows that F^* is uniquely determined by its moments and using DCT we get that $\beta^*_{h,d} \to \beta^*_h$. From which it also follows that $F^*_d \xrightarrow{w} F^*$. Proof of part (c) is similar to that of Theorem 1 (c). □

Remark 7. *We have not proved Theorem 2 under the condition of* $\{\varepsilon_t\}$ *being i.i.d with finite second moment. This is because there is now no straightforward way to apply (3.1). We cannot apply (3.2) either since* $\Gamma_n^*(X)$ *is not nonnegative definite. Simulation results indicate that the same LSD continues to hold for the i.i.d. finite second moment case. We conjecture that the LSDs* F_d^* *and* F^* *exist under this condition. Moreover,* F_d^* *should converge weakly to* F^* *under this condition.*

4 Some properties of the limiting spectral distributions

It appears to be hard to find out $p_w^{(d)}$ for every w or even find out the limit moments. In this section we provide a miscellany of properties on these. In particular, we give several sufficient conditions for $p_w^{(d)}$ to be zero in Lemma 6, Lemma 7 and Lemma 8. Lemma 7 provides a result which is parallel to a similar result proved by Bryc Dembo and Jiang (2006) that had turned out to be a crucial element in the proof of their main theorem. The last Lemma (Lemma 9) is used in the proof of Theorem 5. Fix any d and a finite partition $\mathcal{P} = \cup_k \{i_k, j_k\}, \ i_k < j_k$. Let

$$\mathcal{R} = \{i \in S_3(w) : \quad w[i] \text{ matches with } \ w[j+h], \ i \neq j\}.$$

Lemma 6. *Suppose* $\mathcal{P}$ *is such that* $S_5 = \phi$. *If* $\mathcal{R} \neq \phi$ *then for any* $w \in \mathcal{P}$, $p_w^{(d)} = 0$.

Proof. Note that $S_5(w) = \phi$ automatically implies $S_6(w) = \phi$. Thus

$$\#S_3(w) = \#S_4(w) = h$$

which in turn implies

$$S_1(w) = S_2(w) = \phi.$$

Hence whenever $w[i_1] \approx w[i_2]$ we must have $i_1 \in S_3(w)$ and $i_2 \in S_4(w)$.

Now note that if $i \in \mathcal{R}$ and if $w[i] \approx w[j+h]$, then $j \in \mathcal{R}$. This implies that we can write $\mathcal{R}$ as a disjoint union of sets $\mathcal{S}_i = \{j_{k_{n_i+1}}, \ldots, j_{k_{n_{i+1}}}\}, \ \sum n_i = \#\mathcal{R}$ such that any relation between $S_3(w)$ and $S_4(w)$ can be written as

$$t_{j_{k_{n_i+p}}} = t_{j_{k_{n_i+p+1}}} + |\pi_{j_{k_{n_i+p+1}}-1} - \pi_{j_{k_{n_i+p+1}}}| + n_{j_{k_{n_i+p}}} \tag{4.1}$$

for some i and $p = 1, \ldots, n_{i+1} - n_i$ (we take $j_{k_{n_{i+1}+1}} = j_{k_{n_i+1}}$ in the above relation).

Here is an example. Fix $d = 1$ and consider the word $w^* = w_0^1 w_0^2 w_0^3 w_0^4 w_{n_5}^4 w_{n_6}^3 w_{n_7}^2 w_{n_8}^1$. Then the relation set (4.1) turns out to be

$$\begin{aligned} t_1 &= t_4 + |\pi_3 - \pi_4| + n_8 \\ t_2 &= t_3 + |\pi_2 - \pi_3| + n_7 \\ t_3 &= t_2 + |\pi_1 - \pi_2| + n_6 \\ t_4 &= t_1 + |\pi_0 - \pi_1| + n_5. \end{aligned} \tag{4.2}$$

So here $\mathcal{R} = \{1, 2, 3, 4\}$, $\mathcal{S}_1 = \{1, 4\}$ and $\mathcal{S}_2 = \{2, 3\}$.

Now taking sum over $p = 1, \ldots, n_{i+1} - n_i - 1$ in (4.1) we get

$$t_{j_{k_{n_i+1}}} - t_{j_{k_{n_{i+1}}}} \geq \sum_{p=1}^{n_{i+1}-n_i-1} n_{j_{k_{n_i+p}}}$$

and for $p = n_{i+1} - n_i$ from (4.1) we get

$$t_{j_{k_{n_{i+1}}}} - t_{j_{k_{n_i+1}}} \geq n_{j_{k_{n_{i+1}}}}.$$

Combining the above two expressions we get that for some C_1 and C_2 which are functions of n_j's,

$$t_{j_{k_{n_i+1}}} + C_1 \leq t_{j_{k_{n_{i+1}}}} \leq t_{j_{k_{n_i+1}}} + C_2. \tag{4.3}$$

Note that $t_i \in G$ for all i and $|C_1|, |C_2| \leq dh$. Thus if we integrate over the region described by (4.3) the limiting contribution will be zero. To illustrate, consider the earlier example again. Then we have

$$t_1 - t_4 \geq n_8 \;\text{ and }\; t_4 - t_1 \geq n_5.$$

Thus combining the two we have

$$t_4 + n_8 \leq t_1 \leq t_4 - n_5.$$

Thus

$$p_w(d) \leq \lim_n \mathbb{E}[\mathbb{I}(t_4/n + n_8/n \leq t_1/n \leq t_4 - n_5/n)] = \mathbb{E}[I(U_1 = U_2)] = 0$$

where U_1, U_2 are i.i.d. $U(0,\ 1)$ random variables. This completes the proof. $\square$

The next Lemma shows that if we have the relation $t_i + |\pi_{i-1} - \pi_i| = t_j + |\pi_{j-1} - \pi_j| + n_{j+h}$, then only those words contribute for which $(\pi_{i-1} - \pi_i)$ and $(\pi_{j-1} - \pi_j)$ are of opposite sign. This is identical to what happens for the Toeplitz matrix with i.i.d. input entries (see Bryc, Dembo and Jiang (2006)). Let

$$\mathcal{M} = \{i : \; w[i+h] \text{ matches with } w[j+h],\; i+h \in S_5(w), j+h \in S_6(w) \;\text{ and }\; b_i b_j = 1\}.$$

Lemma 7. *If $\mathcal{M} \neq \phi$, then for any $w \in \mathcal{P}$, $p_w^{(d)} = 0$. As a consequence $\{p_{\mathbf{k}}^{(d)}\} = 0$.*

Proof. Consider i, j such that $w[i+h]$ matches with $w[j+h]$, $i+h \in S_5(w)$, $j+h \in S_6(w)$. Then if $b_i = b_j$ then from (3.21) we get

$$(\pi_{i-1} - \pi_i) - (\pi_{j-1} - \pi_j) = b_j(t_j - t_i) + b_j n_{j+h}. \tag{4.4}$$

On the other hand, if $b_i = -b_j$,

$$(\pi_{i-1} - \pi_i) + (\pi_{j-1} - \pi_j) = b_j(t_i - t_j) - b_j n_{j+h}. \tag{4.5}$$

Now consider i, j such that $w[i]$ matches with $w[j+h]$, $i \in S_3(w)$, $j+h \in S_4(w)$. Then from (3.20),

$$\pi_{j-1} - \pi_j = b_j(t_i - t_j) - b_j n_{j+h}. \tag{4.6}$$

Now using (4.4)–(4.6) we get

$$\begin{aligned}
\pi_0 - \pi_h &= \sum_{i=1}^{h} (\pi_{i-1} - \pi_i) \\
&= \sum_{j+h \in S_4(w)} (\pi_{j-1} - \pi_j) + \sum_{i+h \in S_5(w)} (\pi_{i-1} - \pi_i) + \sum_{j+h \in S_6(w)} (\pi_{j-1} - \pi_j) \\
&= \sum_{\substack{w[i] \approx w[j+h] \\ i \in S_3(w)}} b_j(t_i - t_j) - b_j n_{j+h} + \sum_{\substack{w[i+h] \approx w[j+h] \\ i+h \in S_5(w) \cap \mathcal{M}}} 2(\pi_{i-1} - \pi_i) + b_j(t_i - t_j) - b_j n_{j+h} \\
&\quad + \sum_{\substack{w[i+h] \approx w[j+h] \\ i+h \in S_5(w) \cap \mathcal{M}^c}} b_j(t_i - t_j) - b_j n_{j+h}. \qquad (4.7)
\end{aligned}$$

The right side of (4.7) is a linear combination of π_i's and t_i's and by writing all nongenerating vertices in terms of generating vertices, we can assume that the above is a linear combination of only generating vertices.

Now we show that if $\mathcal{M} \neq \phi$ then there exists at least one generating vertex such that its coefficient in the above is nonzero. Let i^* be the largest index such that $i + h \in S_5(w)$. Note that $\{\lambda_{\mathbf{j}}^{\mathbf{b}}\}$ depend only on the vertices which are to the left of it. Thus from (4.7) it is clear that the coefficient of π_{i^*} is nonzero and hence it is a generating vertex. Now note that

$$\pi_0 - \pi_h = \lambda_{\mathbf{h+1}}^{\mathbf{b}} - \lambda_{\mathbf{2h+1}}^{\mathbf{b}} + m_{h+1} - m_{2h+1}.$$

Since coefficient of π_{i^*} is not zero, we cannot have $\lambda_{\mathbf{h+1}}^{\mathbf{b}} - \lambda_{\mathbf{2h+1}}^{\mathbf{b}} \equiv 0$. Hence $p_w^{(d)} = 0$. □

Let us now investigate what happens if $\mathcal{M} = \phi$. This development will be used in the next two Lemmata. This implies

$$\pi_0 - \pi_h = \sum_{\substack{w[i] \approx w[j+h] \\ i \in S_3(w)}} b_j(t_i - t_j) - b_j n_{j+h} + \sum_{\substack{w[i+h] \approx w[j+h] \\ i+h \in S_5(w)}} b_j(t_i - t_j) - b_j n_{j+h}. \qquad (4.8)$$

From our construction of $\lambda_{\mathbf{j}}^{\mathbf{b}}$ it follows that if $d = 0$ then the only changes in (4.8) will be that $n_j = 0$ for all j nongenerating vertices. Thus having $\lambda_{\mathbf{h+1}}^{\mathbf{b}} - \lambda_{\mathbf{2h+1}}^{\mathbf{b}} \equiv 0$ (for $d = 0$) will imply

$$\sum_{\substack{w[i] \approx w[j+h] \\ i \in S_3(w)}} b_j(t_i - t_j) + \sum_{\substack{w[i+h] \approx w[j+h] \\ i+h \in S_5(w)}} b_j(t_i - t_j) \equiv 0 \qquad (4.9)$$

and for $d > 0$ the following will hold

$$\sum_{\substack{w[i] \approx w[j+h] \\ i \in S_3(w)}} b_j(\lambda_{\mathbf{i}}^{\mathbf{b}} - \lambda_{\mathbf{j}}^{\mathbf{b}}) + \sum_{\substack{w[i+h] \approx w[j+h] \\ i+h \in S_5(w)}} b_j(\lambda_{\mathbf{i}}^{\mathbf{b}} - \lambda_{\mathbf{j}}^{\mathbf{b}}) \equiv 0. \qquad (4.10)$$

Lemma 8. *Suppose* $\lambda_{\mathbf{h+1}}^{\mathbf{b}} - \lambda_{\mathbf{2h+1}}^{\mathbf{b}} \equiv 0$. *Then* $m_{2h+1} = \sum c_j n_j$ *where* $|c_j| = 1$ *for all nongenerating vertices* j.

Proof. Since $b_i \in \{-1, 1\}$, once we express m_{2h+1} as a sum of $c_j n_j$, so whenever the coefficient is nonzero, it has absolute value one. Note that if (i) $i \in S_3(w)$, $w[i] \approx w[j+h]$ or if (ii) $i+h \in S_5(w)$, $w[i+h] \approx w[j+h]$ then either $j \in S_1(w)$ or $j \in S_2(w)$ and for $i+h \in S_5(w)$, $w[i+h] \approx w[j+h]$, either $i \in S_1(w)$ or $i \in S_2(w)$. If $j \in S_1(w)$ then we $t_j = \lambda^{\mathbf{b}}_{\mathbf{j}}$ and if $j \in S_2(w)$ we have $t_j = \lambda^{\mathbf{b}}_{\mathbf{j}} + n_j$. Thus from (4.8) and (4.10) we have

$$\begin{aligned}
\pi_0 - \pi_h &= \sum_{\substack{w[i]\approx w[j+h] \\ i\in S_3(w)}} b_j(\lambda^{\mathbf{b}}_{\mathbf{i}} - \lambda^{\mathbf{b}}_{\mathbf{j}}) - b_j n_{j+h} + \sum_{\substack{w[i+h]\approx w[j+h] \\ i+h\in S_5(w)}} b_j(\lambda^{\mathbf{b}}_{\mathbf{i}} - \lambda^{\mathbf{b}}_{\mathbf{j}}) - b_j n_{j+h} \\
&\quad - \sum_{\substack{w[i]\approx w[j+h] \\ i\in S_3(w), j\in S_2(w)}} b_j n_j - \sum_{\substack{w[i+h]\approx w[j+h] \\ i+h\in S_3(w), j\in S_2(w)}} b_j n_j + \sum_{\substack{w[i+h]\approx w[j+h] \\ i+h\in S_3(w), i\in S_2(w)}} b_j n_i \\
&= - \sum_{\substack{w[i]\approx w[j+h] \\ i\in S_3(w)}} b_j n_{j+h} - \sum_{\substack{w[i+h]\approx w[j+h] \\ i+h\in S_5(w)}} b_j n_{j+h} - \sum_{\substack{w[i]\approx w[j+h] \\ i\in S_3(w), j\in S_2(w)}} b_j n_j \\
&\quad - \sum_{\substack{w[i+h]\approx w[j+h] \\ i+h\in S_5(w), j\in S_2(w)}} b_j n_j + \sum_{\substack{w[i+h]\approx w[j+h] \\ i+h\in S_5(w), i\in S_2(w)}} b_j n_i \\
&= m_{h+1} - m_{2h+1} = -m_{2h+1} \ (m_{h+1} = 0 \quad \text{by definition}). \qquad (4.11)
\end{aligned}$$

Thus m_{2h+1} is a linear combination of n_j's and all the coefficients have absolute value one. In (4.11) we have expressed m_{2h+1} as a sum of five different quantities. Now it remains to argue that for any j nongenerating, the coefficient corresponding to j is nonzero.

Fix any j_0 nongenerating. If $j_0 \in S_2(w)$ then $j_0 + h \in S_4(w) \cup S_5(w) \cup S_6(w)$. For $j_0 + h \in S_4(w)$, $j_0 + h \in S_6(w)$ and $j_0 + h \in S_5(w)$, n_{j_0} is present in respectively, the third, the fourth and the last summation terms in (4.11). Likewise, if $j_0 \in S_4(w)$ or $j_0 \in S_6(w)$ then n_{j_0} is present in the first and second summation terms respectively. This shows that whenever j_0 is nongenerating, the coefficient of n_{j_0} is nonzero in (4.11). This completes the proof of the Lemma. □

The following Lemma will be used in the proof of Theorem 5.

Lemma 9. *Fix any d finite and any sequence of nonnegative integers k_i, $i = 0, 1, 2, \ldots, d$ such that $\sum_{i \text{ odd}, \ i\le d} k_i$ is odd. Then $p_{\mathbf{k}} = 0$. As a consequence, when k_1 is odd $p^{(1)}_{k_0,k_1} = 0$ and $p^{(2)}_{k_0,k_1,k_2} = 0$ for any choices of k_0, k_2.*

Proof. Fix any partition $\mathcal{P}$ and consider any word $w \in \mathcal{P} \cap \mathcal{W}(\mathbf{k})$. For $p^{(d)}_w$ to be positive we must have

$$\lambda^{\mathbf{b}}_{\mathbf{h+1}} - \lambda^{\mathbf{b}}_{\mathbf{2h+1}} \equiv 0 \ \text{ and } \ m_{2h+1} = m_{h+1}.$$

Now from Lemma 7 and 8, we know that

$$m_{2h+1} - m_{h+1} = \sum_{j\in S_2(w)\cup S_4(w)\cup S_6(w)} c_j n_j$$

where

$$|c_j| = 1 \ \text{ for all } \ j, \ n_j \in \{-d, \ldots, 0, \ldots, d\}.$$

Let

$$k_i = \#\{j : \ |n_j| = i\}, i = 0, 1, \ldots, d.$$

Note that $m_{2h+1} - m_{h+1} = 0$ implies

$$\sum_{\substack{j \text{ non generating}, \\ c_j = 1}} n_j = \sum_{\substack{j \text{ non generating}, \\ c_j = -1}} n_j$$

which in turn implies that

$$\sum_{j \text{ non generating}} n_j \text{ is even.}$$

But

$$\sum_{i \text{ odd } i \leq d} k_i \text{ is odd.}$$

This implies that

$$\sum_{\substack{j \text{ non generating}, \\ |n_j| \text{ odd}}} n_j \text{ is also odd.}$$

Since for any choices of $k_0, k_2, k_4, \ldots, k_{2\lfloor d/2 \rfloor}$,

$$\sum_{\substack{j \text{ non generating}, \\ |n_j| \text{ even}}} n_j \text{ is even,}$$

we arrive at a contradiction. Hence we must have $m_{2h+1} - m_{h+1} \neq 0$ and thus $p_w^{(d)} = 0$ for any $p \in \mathcal{P} \cap \mathcal{W}(\mathbf{k})$. This completes the proof of the Lemma. $\square$

4.1 Proof of Theorem 5

We shall prove the result only for finite d and the matrix $\Gamma_n(X)$. Same proof works for $\Gamma_n^*(X)$ for finite d. Since, for $d = \infty$, the LSD is the weak limit of the LSDs obtained for d finite, the result continues to hold when d is infinite.

From (2.4) and using Lemma 9 we have

$$\beta_{h,d} = \sum^{*} p_{\mathbf{k}}^{(d)} \prod_{i=0}^{d} [\gamma_X(i)]^{k_i}$$

where $\sum^{*}$ denotes summation over all $k_0, k_1, \ldots, k_d$ nonnegative integer such that

$$k_1 + k_3 + \ldots + k_{2\lfloor (d-1)/2 \rfloor + 1} \text{ is even.} \tag{4.12}$$

Let $\beta_{h,d}^i$, $i = 1, 2, 3$, denote the h^{th} moment of the LSD when X is $MA(d)$ with parameters respectively $(\theta_0, \theta_1, \ldots, \theta_d)$, $(\theta_0, -\theta_1, \ldots, (-1)^i\theta_i, \ldots, (-1)^d\theta_d)$ and $(-\theta_0, \theta_1, \ldots, (-1)^{i+1}\theta_i, \ldots, (-1)^{d+1}\theta_d)$ respectively and denote the corresponding three sequence by T, Y and Z respectively. Note that $p^{(d)}_{\mathbf{k}}$ is same for every choices of $\theta_0, \theta_1, \ldots, \theta_d$ and the LSD here is uniquely determined by its moments. Thus we only need to show that the coefficient of $p^{(d)}_{\mathbf{k}}$ is same for these three choices of parameters whenever (4.12) holds. Now note that for any j,

$$\begin{aligned}\gamma_T(j) &= \sum_{i=0}^{d-j} \theta_i \theta_{i+j} \\ &= (-1)^j \sum_{i=0}^{d-j} (-1)^i \theta_i (-1)^{i+j} \theta_{i+j} = (-1)^j \gamma_Y(j). \qquad (4.13)\end{aligned}$$

Thus

$$\begin{aligned}[\gamma_T(0)]^{k_0} \cdots [\gamma_T(d)]^{k_d} &= \left[\prod_{j=0}^{d} (-1)^{jk_j}\right] \times [\gamma_Y(0)]^{k_0} \cdots [\gamma_{Y^2}(d)]^{k_d} \\ &= (-1)^{k_1+k_3+\ldots+k_{2\lfloor (d-1)/2 \rfloor+1}} \times \gamma_Y(0)^{k_0} \ldots \gamma_Y(d)^{k_d} \\ &= [\gamma_Y(0)]^{k_0} \cdots [\gamma_Y(d)]^{k_d}. \qquad (4.14)\end{aligned}$$

Similarly one can show that whenever (4.12) holds then

$$[\gamma_T(0)]^{k_0} \cdots [\gamma_T(d)]^{k_d} = [\gamma_Z(0)]^{k_0} \cdots [\gamma_Z(d)]^{k_d}.$$

Thus we have $\beta^1_{h,d} = \beta^2_{h,d} = \beta^3_{h,d}$ for every h and the proof is complete. □

References

[1] Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, 9, 611-677 (with discussions).

[2] Bai, Z. D. and Silverstein, J. (2006). *Spectral Analysis of Large Dimensional Random Matrices.* Science Press, Beijing.

[3] Basak, Anirban (2009). *Large dimensional random matrices. M. Stat. Project report*, May 2009. Indian Statistical Institute, Kolkata.

[4] Basak, Anirban and Bose, Arup (2010). Balanced random Toeplitz and Hankel matrices. *Electronic Communications in Probability*, 15, 134-148.

[5] Bhatia, R. (1997). *Matrix Analysis.* Springer, New York.

[6] Bose, Arup and Sen, Arnab (2008). Another look at the moment method for large dimentional random matrices. *Electronic Journal of Probability*, 13, no.21, 588-628.

[7] Böttcher, A. and Silberman, B. (1998) *Introduction to Large Truncated Toeplitz Matrices*, Springer-Verlag, New York.

[8] Bryc, Włodzimierz, Dembo, Amir and Jiang, Tiefeng (2006). Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.*, 34, no. 1, 1–38.

[9] Dudley, Richard M. (2002). *Real analysis and probability.* Cambridge University Press, Cambridge.

[10] Hammond, C. and Miller, S. J. (2005). Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices. *J. Theoret. Probab.* 18, no. 3, 537–566.

[11] Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices, (Russian) *Mat. Sb. (N.S.)* 72 (114), 507–536.

[12] McMurry, Timothy, L. and Politis, Dimitris N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31, 6, 471–482.

[13] Sen, Arnab (2006). *Large dimensional random matrices. M. Stat. Project report*, May 2006. Indian Statistical Institute, Kolkata.

[14] Sen, Sanchayan (2010). *Limiting spectral distribution of random matrices. M.Stat. Project report*, July 2010. Indian Statistical Institute, Kolkata.

[15] Wu, Wei Biao and Pourahmadi, Mohsen (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica* 19, 1755-1768.

[16] Xiao, Han and Wu, Wei Biao (2011). Covariance estimation for stationary time series. Preprint. Dept. of Statistics, Univeristy of Chicago.