

Limiting spectral distribution of sample autocovariance matrices

ANIRBAN BASAK¹, ARUP BOSE² and SANCHAYAN SEN³

¹*Department of Statistics, Stanford University, 390 Serra Mall, Stanford, CA 94305-4065, USA.*

E-mail: anirbanb@stanford.edu

²*Statistics and Mathematics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India.*

E-mail: bosearu@gmail.com

³*Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA. E-mail: sen@cims.nyu.edu*

We show that the empirical spectral distribution (ESD) of the sample autocovariance matrix (ACVM) converges as the dimension increases, when the time series is a linear process with reasonable restriction on the coefficients. The limit does not depend on the distribution of the underlying driving i.i.d. sequence and its support is unbounded. This limit does not coincide with the spectral distribution of the theoretical ACVM. However, it does so if we consider a suitably tapered version of the sample ACVM. For banded sample ACVM the limit has unbounded support as long as the number of non-zero diagonals in proportion to the dimension of the matrix is bounded away from zero. If this ratio tends to zero, then the limit exists and again coincides with the spectral distribution of the theoretical ACVM. Finally, we also study the LSD of a naturally modified version of the ACVM which is not non-negative definite.

Keywords: autocovariance function; autocovariance matrix; banded and tapered autocovariance matrix; linear process; spectral distribution; stationary process; Toeplitz matrix

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* (ACVF) $\gamma_X(\cdot)$ and the *autocovariance matrix* (ACVM) $\Sigma_n(X)$ of order n are defined as:

$$\gamma_X(k) = \text{cov}(X_0, X_k), \quad k = 0, 1, \dots$$

and

$$\Sigma_n(X) = ((\gamma_X(i - j)))_{1 \leq i, j \leq n}.$$

To every ACVF, there corresponds a unique distribution, called the *spectral distribution*, $F_X(\cdot)$ which satisfies

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

We shall assume that

$$\sum_{k=1}^{\infty} |\gamma_X(k)| < \infty. \tag{1.2}$$

Then $F_X(\cdot)$ has a density, known as the *spectral density* of X or of $\gamma_X(\cdot)$, which equals

$$f_X(t) = \sum_{k=-\infty}^{\infty} \exp(-2\pi itk) \gamma_X(k), \quad t \in (0, 1]. \tag{1.3}$$

The *non-negative definite* estimate of $\Sigma_n(X)$ is the *sample ACVM*

$$\Gamma_n(X) = ((\hat{\gamma}_X(i - j)))_{1 \leq i, j \leq n} \quad \text{where } \hat{\gamma}_X(k) = n^{-1} \sum_{i=1}^{n-|k|} X_i X_{i+|k|}. \tag{1.4}$$

The matrix $\Gamma_n(X)$ is a random matrix. Study of the behavior of random matrices, when the dimension goes to ∞ , have been inspired by both theory and applications. This is done by studying the behavior of its eigenvalues. For instance a host of results are known for the related sample covariance matrix, in the i.i.d. set-up and its variations; results on its spectral distribution, spacings of the eigenvalues, spectral statistics etc. encompasses a rich theory and a variety of applications.

The autocovariances are of course crucial objects in time series analysis. They are used in estimation, prediction, model fitting and white noise tests. Under suitable assumptions on $\{X_t\}$, for every fixed k , $\hat{\gamma}_X(k) \rightarrow \gamma_X(k)$ almost surely (a.s.). There are also results on the asymptotic distribution of specific functionals of the autocovariances. Recently, there has been growing interest in the matrix $\Gamma_n(X)$ itself. For instance, the largest eigenvalue of $\Sigma_n(X) - \Gamma_n(X)$ does not converge to zero, even under reasonable assumptions (see Wu and Pourahmadi [17], Arcones [14] and Xiao and Wu [18]).

In this article we study the behavior of $\Gamma_n(X)$, and a few other natural estimators of $\Sigma_n(X)$, as $n \rightarrow \infty$, through the behavior of its spectral distribution. We investigate the *consistency* (in an appropriate sense) of these estimators.

For a real symmetric matrix $A_{n \times n}$ with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ the *Empirical Spectral Distribution (ESD)* of A_n is defined as,

$$F^{A_n}(x) = n^{-1} \sum_{i=1}^n \mathbb{I}(\lambda_i \leq x). \tag{1.5}$$

If $\{F^{A_n}\}$ converges weakly to F , we write $F^{A_n} \xrightarrow{w} F$. For X any random variable with distribution F , X or F will be called the *Limiting Spectral Distribution (or measure) (LSD)* of F^{A_n} . The entries of A_n are allowed to be random. In that case, the limit is taken to be either in probability or (as in this paper) in a.s. sense.

Any matrix T_n of the form $((t_{i-j}))_{1 \leq i, j \leq n}$ is a *Toeplitz* matrix and hence $\Sigma_n(X)$ and $\Gamma_n(X)$ (with a triangular sequence of entries) are Toeplitz matrices. For T_n symmetric, from Szegő's theory of Toeplitz operators (see Böttcher and Silbermann [9]), we note that if $\sum |t_k| < \infty$, then the LSD of T_n equals $f(U)$ where U is uniformly distributed on $(0, 1]$ and $f(x) =$

$\sum_{k=-\infty}^{\infty} t_k \exp(-2\pi i x k)$, $x \in (0, 1]$. In particular if (1.2) holds, then the LSD of $\Sigma_n(X)$ equals $f_X(U)$ where $f_X(\cdot)$ is as defined in (1.3).

We call a sequence of estimators $\{E_n\}$ of $\Sigma_n(X)$ *consistent* if its LSD is $f_X(U)$ where U is uniformly distributed on $[0, 1]$. We show that $\{\Gamma_n(X)\}$ is inconsistent (see Theorem 2.1(c)). We also show that if $\Gamma_n(X)$ is modified by suitable tapering or banding then the modified estimators are indeed consistent (see Theorem 2.3(b) and (c)). This phenomenon is mainly due to the estimation of a large number of autocovariances by $\Gamma_n(X)$. Such inconsistency of sample covariance matrices has also been observed in the context of high-dimensional multivariate analysis, and is now well understood, with the help the results from Random Matrix Theory.

To obtain the convergence of ESD of such estimators, we impose a reasonable condition on the stationary process $\{X_t\}$; we assume it to be a linear process, that is,

$$X_t = \sum_{k=0}^{\infty} \theta_k \varepsilon_{t-k}, \quad (1.6)$$

where $\{\theta_k\}$ satisfies a weak condition and $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of independent random variables with appropriate conditions. The simulations of Sen [15] suggested that the LSD of $\Gamma_n(X)$ exists and is independent of the distribution of $\{\varepsilon_t\}$ as long as they are i.i.d. with mean zero and variance one. Basak [4] and Sen [16] initially studied, respectively, the special cases where X is an i.i.d. process or is an MA(1) process.

In Theorem 2.1, we prove that, if $\{X_t\}$ satisfies (1.6) and $\sum_{k=0}^{\infty} |\theta_k| < \infty$ then the LSD of $\Gamma_n(X)$ exists, and it is *universal* when $\{\varepsilon_t\}$ are independent with mean zero and variance 1 and are either uniformly bounded or identically distributed. We further show that LSD is unbounded when $\theta_i \geq 0$ for all i , and thus $\{\Gamma_n(X)\}$ is inconsistent, since $f_X(U)$ is of bounded support.

When $\{X_t\}$ is a finite order process, the limit moments can be written as multinomial type sums of the autocovariances (see (2.4)). When X is of infinite order, the limit moments are the limits of these sums as the order tends to infinity. Additional properties of the limit moments are available in the companion report Basak, Bose and Sen [5].

Incidentally, $\Gamma_n(X)$ reminds us of the sample covariance matrix, S , for the i.i.d. set-up, whose spectral properties are well known. See Bai [3] for the basic references on S . In particular, the LSD of S (with i.i.d. entries) under suitable conditions is the Marčenko–Pastur law and is supported on the interval $[0, 4]$. Thus, the LSD of $\Gamma_n(X)$ is in sharp contrast.

The proof of Theorem 2.1 is challenging, mainly because of the non-linear dependence, and the Toeplitz structure of $\Gamma_n(X)$. Bai and Zhou [2] and Yao [19] study the LSD of the sample covariance matrix of $\mathbf{X}_1, \dots, \mathbf{X}_n$ where \mathbf{X}_k are i.i.d. p -dimensional vectors with some dependence structure. They establish the existence of the LSD by using Stieltjes transform method. Here this approach fails completely due to the strong row column dependence. In fact no Stieltjes transform proof for even the Toeplitz matrix with i.i.d. input is known. Moreover one added advantage in both the above articles is the existence of n independent columns, which we lack here, because we have only one sample from the linear process $\{X_t\}$. The methods of Xiao and Wu [18] is also not applicable in our set-up because they deal with only the maximum eigenvalue of the difference of $\Sigma_n(X)$, and $\Gamma_n(X)$, not the ESD of $\Gamma_n(X)$.

Now consider a sequence of integers $m := m_n \rightarrow \infty$, and a kernel function $K(\cdot)$. Define

$$\hat{f}_X(t) = \sum_{k=-m}^m K(k/m) \exp(-2\pi i t k) \hat{\gamma}_X(k), \quad t \in (0, 1] \tag{1.7}$$

as the kernel density estimate of $f_X(\cdot)$. Considering this as a spectral density, the corresponding ACVF is given by (for $-m \leq h \leq m$):

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

and is 0 otherwise. This motivates the consideration of the *tapered sample ACVM*

$$\Gamma_{n,K}(X) = ((K((i - j)/m) \hat{\gamma}_X(i - j)))_{1 \leq i, j \leq n}. \tag{1.8}$$

If K is a non-negative definite function then $\Gamma_{n,K}(X)$ is also non-negative definite. Among other results, Xiao and Wu [18] also showed that under the growth condition $m_n = o(n^\gamma)$ for a suitable γ and suitable conditions on K , the largest eigenvalue of $\Gamma_{n,K}(X) - \Sigma_n(X)$ tends to zero a.s. Theorem 2.3(c) states that under the minimal condition $m_n/n \rightarrow 0$, if K is bounded, symmetric and continuous at 0 and $K(0) = 1$, then $\Gamma_{n,K}(X)$ is consistent. This is a reflection of the fact that the consistency notion of Xiao and Wu [18] in terms of the maximum eigenvalue is stronger than our notion and hence our consistency holds under weaker growth condition on m_n .

The second approach is to use banding as in McMurphy and Politis [14] who used it to develop their bootstrap procedures. We study two such banded matrices. Let $\{m_n\}_{n \in \mathbb{N}} \rightarrow \infty$ be such that $\alpha_n := m_n/n \rightarrow \alpha \in [0, 1]$. 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 $\Gamma_{n,K}$ with $K(x) = I_{\{|x| \leq 1\}}$. The *type II banded ACVM* $\Gamma_n^{\alpha, II}(X)$ is the $m_n \times m_n$ principal sub matrix of $\Gamma_n(X)$. Theorem 2.3(a) and (b) states our results on these banded ACVMs. In particular, the LSD exists for all α and is unbounded when $\alpha \neq 0$. When $\alpha = 0$, the LSD is $f_X(U)$ and thus those estimate matrices are consistent.

A related matrix, which may be of interest, especially to probabilists, is,

$$\Gamma_n^*(X) = ((\gamma_X^*(|i - j|)))_{1 \leq i, j \leq n} \quad \text{where } \gamma_X^*(k) = n^{-1} \sum_{i=1}^n X_i X_{i+k}, k = 0, 1, \dots \tag{1.9}$$

$\Gamma_n^*(X)$ does not have a “data” interpretation unless one assumes we have $2n - 1$ observations X_1, \dots, X_{2n-1} . It is not non-negative definite and hence many of the techniques applied to $\Gamma_n(X)$ are not available for it. Theorem 2.2 states that its LSD also exists but under stricter conditions on $\{X_t\}$. Its moments dominate those of the LSD of $\Gamma_n(X)$ when $\theta_i \geq 0$ for all i (see Theorem 2.2(c)) even though simulations show that the LSD of $\Gamma_n^*(X)$ has significant positive mass on the negative axis.

2. Main results

We shall assume that $X = \{X_t\}_{t \in \mathbb{Z}}$ is a linear (MA(∞)) process

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

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

$$X^{(d)} = \{X_{t,d} \equiv \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_d \varepsilon_{t-d}, t \in \mathbb{Z}\} \quad (\theta_0 \neq 0).$$

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

The series in (2.1) converges a.s. under Assumptions A(a) (or (b)) and B(b). Further, X and $X^{(d)}$ are strongly stationary and ergodic under Assumption A(a) and weakly (second order) stationary under Assumptions A(b) and B(b).

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

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

Let $\{k_i\}$ stand for suitable integers and let

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

Theorem 2.1 (Sample ACVM). Suppose Assumption A(a) or (b) holds.

(a) Then a.s., $F^{\Gamma_n(X^{(d)})} \xrightarrow{w} F_d$ which is non-random and does not depend on the distribution of $\{\varepsilon_t\}$. Further,

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

where $\{p_{\mathbf{k}}^{(d)}\}$ are universal constants independent of the θ_i and the $\{\varepsilon_t\}$. They are defined by a limiting process given in (3.11) and (3.25).

(b) Under Assumption B(b), a.s., $F^{\Gamma_n(X)} \xrightarrow{w} F$ which is non-random and independent of the distribution of $\{\varepsilon_t\}$. Further for every fixed h , as $d \rightarrow \infty$,

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

(c) Under Assumption B(a), F_d has unbounded support and $\beta_{h,d-1} \leq \beta_{h,d}$ if $d \geq 1$. Consequently, if Assumption B(a) and (b) holds, then F has unbounded support. Therefore $\{\Gamma_n(X)\}$ is inconsistent.

Theorem 2.2. Suppose Assumption A(b) holds. Then conclusions of Theorem 2.1 continue to hold for $\Gamma_n^*(X)$, $d \leq \infty$, and (2.4) holds with modified universal constants $\{p_k^{*(d)}\}$.

Remark 2.1. (i) From the proofs, it will follow that the limit moments $\{\beta_{h,d}\}$ and $\{\beta_h\}$ of the above LSDs are dominated by $\frac{4^h(2h)!}{h!} (\sum_{k=0}^\infty |\theta_k|)^{2h}$ which are the $(2h)$ th moment of a Gaussian variable with mean zero and variance $4((\sum_{k=0}^\infty |\theta_k|)^2)$. Hence the limit moments uniquely identify the LSDs.

(ii) All the above LSDs have unbounded support while $f_X(U)$ has support contained in $[-\sum_{-\infty}^\infty |\gamma_X(k)|, \sum_{-\infty}^\infty |\gamma_X(k)|]$. Simulations show that the LSD of $\Gamma_n^*(X)$ has positive mass on the negative real axis.

(iii) Since $\Gamma_n^*(X)$ is not non-negative definite, the proof of Theorem 2.2 for $d = \infty$ is different from the proof of Theorem 2.1 and needs Assumption A(b). A detailed discussion on the different assumptions is given in Remark 3.1 at the end of the proofs.

(iv) Unfortunately, the moments of the LSD of $\Gamma_n(X)$ has no easy description. There is no easy description of the constants $\{p_k^{(d)}\}$ either. To explain briefly the complications involved in providing explicit expressions for these quantities, consider the much simpler random Toeplitz matrix $n^{-1/2}T_{n,\varepsilon} = n^{-1/2}((\varepsilon_{|i-j|}))$ where $\{\varepsilon_t\}$ is i.i.d. with mean zero variance 1. Bryc, Dembo and Jiang [10] and Hammond and Miller [13] have showed that the LSD exists and is universal. The limit moments are of the form

$$\beta_{2k}(T) = \sum p(w),$$

where the sum is over the so called matched words w and for each w , $p(w)$ is given as the volume of a suitable subset of a k -dimensional hypercube. These subsets are defined through the intersection of k hyperplanes which arise from the function $L(i, j) = |i - j|$. Thus the value of $p(w)$ can be calculated by performing multiple integration but must be done only via numerical integration when k becomes large. For more details, see Bose and Sen [8]. For our set up, definition of matched words is generalised and is given in Section 3 and $p_k^{(d)}$ are given by more complicated integrals. This is the main reason why the moments of the LSD cannot be obtained in any closed form, even when X is the i.i.d. process.

Bose and Sen [8] considered the Toeplitz matrix $T_{n,X} = ((X_{|i-j|}))$ and showed that its LSD exists under suitable conditions. The moments β_{2k}^* of the LSD can be written in terms of $\{\theta_j\}$

and $\{\beta_{2k}(T)\}$. This relation is given by

$$\beta_{2k}^* = \mathbb{E} \left| \sum_{j=0}^{\infty} \theta_j \exp(-2\pi i j U) \right|^{2k} \beta_{2k}(T), \tag{2.5}$$

where U is uniformly distributed on $(0, 1)$.

Even a relation like (2.5) relating the i.i.d. process case to the linear process case eludes us for the autocovariance matrix. This is primarily due to the non-linear dependence of the autocovariances $\{\hat{\gamma}_X(k)\}$ on the driving $\{\varepsilon_t\}$. One of the Referees has pointed out that in this context, the so called ‘‘diagram formula’’ (see Arcones [1], Giraitis, Robinson and Surgailis [12] for details) may be useful, presumably to obtain a formula relating the linear process case to the i.i.d. case.

It is also noteworthy that no limit moment formula or explicit description of the LSD is known for the matrix $n^{-1}H_{n,\varepsilon}H'_{n,\varepsilon}$ where $H_{n,\varepsilon}$ is the *non-symmetric* Toeplitz matrix defined using an i.i.d. sequence (see Bose, Gangopadhyay and Sen [7]).

Theorem 2.3 (Banded and tapered sample ACVM). *Suppose Assumption A(b) holds.*

(a) *Let $0 < \alpha \leq 1$. Then all the conclusions of Theorem 2.1 hold for $\Gamma_n^{\alpha,I}(X^{(d)})$ and $\Gamma_n^{\alpha,II}(X^{(d)})$ with modified universal constants $\{p_k^{\alpha,I,(d)}\}$ and $\{p_k^{\alpha,II,(d)}\}$, respectively, in (2.4). Same conclusions continue to hold also for $d = \infty$.*

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

(a) and (b) *remain true for $\Gamma_n^{\alpha,II}(X^{(d)})$ and $\Gamma_n^{\alpha,II}(X)$ under Assumption A(a).*

(c) *Suppose Assumption B(b) holds. Let K be bounded, symmetric and continuous at 0, $K(0) = 1$, $K(x) = 0$ for $|x| > 1$. Suppose $m_n \rightarrow \infty$ such that $m_n/n \rightarrow 0$. Then the LSD of $\Gamma_{n,K}(X)$ is $f_X(U)$ for $d \leq \infty$.*

Remark 2.2. (i) When K is non-negative definite, Theorem 2.3(c) holds under Assumption A(a).

(ii) Xiao and Wu [18] 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 a.s.

(iii) Each of the LSDs above are identical for the combinations $(\theta_0, \theta_1, \theta_2, \dots)$, $(\theta_0, -\theta_1, \theta_2, \dots)$ and $(-\theta_0, \theta_1, -\theta_2, \dots)$. See Basak, Bose and Sen [5] for a proof which is based on properties of the limit moments. The LSDs $f_X(U)$ of $\Sigma_n(X)$ are identical for processes with autocovariances $(\gamma_0, \gamma_1, \dots, \gamma_d)$ and $(\gamma_0, -\gamma_1, \dots, (-1)^d \gamma_d)$. The same is true of all the above LSDs.

3. Proofs

Szegö’s theorem (or its triangular version) for non-random Toeplitz matrices needs summability (or square summability) of the entries and that is absent (in the a.s. sense) for $\Gamma_n(X)$. As an answer to a question raised by Bai [3], Bryc, Dembo and Jiang [10] and Hammond and Miller [13] showed that for the random Toeplitz matrix $n^{-1/2}T_{n,\varepsilon} = n^{-1/2}((\varepsilon_{|i-j|}))$ where $\{\varepsilon_t\}$ is i.i.d. with mean zero variance 1, the LSD exists and is universal (does not depend on the underlying distribution of ε_1). Bose and Sen [8] considered the Toeplitz matrix $T_{n,X} = ((X_{|i-j|}))$ and showed that

the LSD of $n^{-1/2}T_{n,X}$ exists under the following condition: X satisfies (1.6), $\sum_{j=0}^{\infty} |\theta_j| < \infty$; further, $\{\varepsilon_j\}$ are independent with mean zero and variance 1 and are (i) either uniformly bounded or (ii) are identically distributed and $\sum_{j=0}^{\infty} j\theta_j^2 < \infty$. However, none of the above two results are applicable to $\Gamma_n(X)$ due to the non-linear dependence of $\hat{\gamma}_X(k)$ on $\{X_t\}$.

Our two main tools will be (i) the moment method to show convergence of distribution and (ii) the bounded Lipschitz metric to reduce the unbounded case to the bounded case and also to prove the results for the infinite order case from the finite order case. 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:

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

Then the LSD of $\{A_n\}$ exists a.s. and is uniquely identified by its moments $\{\beta_h\}$ given below if the following three conditions hold:

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

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

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

Let d_{BL} denote the *bounded Lipschitz metric* on the space of probability measures on \mathbb{R} , topologising the weak convergence of probability measures (see Dudley [11]). The following lemma and its proof is given in Bai [3].

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

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

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

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

When $\alpha = 1$, then without loss of generality for asymptotic purposes, we assume that $m_n = n$. We visualise the full ACVM $\Gamma_n(X)$ as the case with $\alpha = 1$. When $\{X_t\}$ is a finite order moving average process with bounded $\{\varepsilon_t\}$, we use the *method of moments* to establish Theorem 2.1(a). The longest and hardest part of the proof is to verify (C1). We first develop a manageable expression for the moments of the ESD and then show that asymptotically only “matched” terms survive. These moments are then written as an iterated sum, where one summation is over finitely many terms (called “words”). Then we verify (C1) by showing that each one of these finitely many terms has a limit. The d_{BL} metric is used to remove the boundedness assumption as well as to deal with the infinite order case. Easy modifications of these arguments yield the existence of the LSD when $0 \leq \alpha \leq 1$ in Theorem 2.3(a) and (b). The proof of Theorem 2.2 is a byproduct of the arguments in the proof of Theorem 2.1. However, due to the matrix now not being non-negative definite, we impose Assumption A(b). The proof of Theorem 2.1(a) is given in details. All other proofs are sketched and details are available in Basak, Bose and Sen [5].

3.1. Proof of Theorem 2.1

The first step is to show that we can without loss of generality, assume that $\{\varepsilon_t\}$ are uniformly bounded so that we can use the moment method. For a standard proof of the following lemma, see Basak, Bose and Sen [5]. For convenience, we will write

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

Lemma 2. *If for every $\{\varepsilon_t\}$ satisfying Assumption A(b), $\Gamma_n(X^{(d)})$ has the same LSD a.s., then this LSD continues to hold if $\{\varepsilon_t\}$ satisfies Assumption A(a).*

Thus from now on we assume that Assumption A(b) holds. Fix any arbitrary positive integer h and consider the h th moment. Then

$$\begin{aligned} \Gamma_{n,d} &= \frac{1}{n} \left((Y_{i,j}^{(n)}) \right)_{i,j=1,\dots,n} \quad \text{where } Y_{i,j}^{(n)} = \sum_{t=1}^n X_{t,d} X_{t+|i-j|,d} \mathbb{I}_{(t+|i-j| \leq n)}, \\ \beta_h(\Gamma_{n,d}) &= \frac{1}{n} \text{Tr}(\Gamma_{n,d}^h) = \frac{1}{n^{h+1}} \sum_{1 \leq \pi_0 = \pi_h, \pi_1, \dots, \pi_{h-1} \leq n} Y_{\pi_0, \pi_1}^{(n)} \cdots Y_{\pi_{h-1}, \pi_h}^{(n)} \\ &= \frac{1}{n^{h+1}} \sum_{\substack{1 \leq \pi_0, \dots, \pi_h \leq n \\ \pi_h = \pi_0}} \left[\prod_{j=1}^h \left(\sum_{t_j=1}^n X_{t_j,d} X_{t_j+|\pi_{j-1}-\pi_j|,d} \mathbb{I}_{(t_j+|\pi_{j-1}-\pi_j| \leq n)} \right) \right]. \end{aligned} \tag{3.3}$$

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

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

Then using (3.3) we can write the so called *trace formula*,

$$\mathbb{E}[\beta_h(\Gamma_{n,d})] = \frac{1}{n^{h+1}} \mathbb{E} \left[\sum_{(\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{A}} X_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \mathbb{I}_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \right]. \tag{3.4}$$

3.1.1. Matching and negligibility of certain terms

By independence of $\{\varepsilon_t\}$, $\mathbb{E}[X_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})}] = 0$ 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 matching and show that certain higher order terms can be asymptotically neglected in (3.4). We say:

- \mathbf{a} is *d-matched* (in short *matched*) if $\forall i \leq 2h, \exists j \neq i$ such that $|a_i - a_j| \leq d$. 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 $\{1, \dots, 2h\}$,

$$\{1, \dots, 2h\} = \bigcup_{k=1}^h \{i_k, j_k\}, \quad i_k < j_k \tag{3.5}$$

such that $\{i_k\}$ are in ascending order and

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

For example, for $d = 1, h = 3$ $(1, 2, 3, 8, 9, 10)$ is matched but not minimal matched and $(1, 2, 5, 6, 9, 10)$ is both matched and minimal matched.

Lemma 3. $\#\{\mathbf{a}: \mathbf{a} \text{ is matched but not minimal matched}\} = O(n^{h-1})$.

Proof. Consider the graph with vertices $\{1, 2, \dots, 2h\}$. Vertices i and j have an edge if $|a_i - a_j| \leq d$. Let $k = \#$ connected components. Consider a typical \mathbf{a} . Let l_j be the number of vertices in the j th component. Since \mathbf{a} is matched, $l_j \geq 2$ for all j and $l_j > 2$ for at least one j . Hence, $2h = \sum_{j=1}^k l_j > 2k$. That implies $k \leq h - 1$. Also if i and j are in the same connected component then $|a_i - a_j| \leq 2dh$. Hence, the number of a_i 's such that i belongs to any given component is $O(n)$ and the result follows. \square

Now we can rewrite (3.4) as

$$\begin{aligned} \mathbb{E}[\beta_h(\Gamma_{n,d})] &= \frac{1}{n^{h+1}} \mathbb{E} \left[\sum_1 X_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \mathbb{I}_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \right] + \frac{1}{n^{h+1}} \mathbb{E} \left[\sum_2 X_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \mathbb{I}_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \right] \\ &\quad + \frac{1}{n^{h+1}} \mathbb{E} \left[\sum_3 X_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \mathbb{I}_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \right] = T_1 + T_2 + T_3 \quad (\text{say}), \end{aligned}$$

where the three summations are over $(\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{A}$ such that $\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})$ 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 Lemma 3, $T_2 \leq \frac{C}{n}$ for some constant C . So provided the limit exists,

$$\lim_{n \rightarrow \infty} \mathbb{E}[\beta_h(\Gamma_{n,d})] = \lim_{n \rightarrow \infty} \frac{1}{n^{h+1}} \mathbb{E} \left[\sum_{\substack{(\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{A}: \\ \mathbf{a}(\mathbf{t}, \boldsymbol{\pi}) \text{ is} \\ \text{minimal matched}}} X_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \mathbb{I}_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} \right]. \tag{3.6}$$

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

3.1.2. Verification of (C1) for Theorem 2.1(a)

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 for $d = 0$ are developed in parallel since this helps to relate the limits in the two cases.

Our starting point is equation (3.6). We first define an *equivalence relation* on the set of minimal matched $\mathbf{a} = \mathbf{a}(\mathbf{t}, \boldsymbol{\pi})$. This yields finitely many equivalence classes. Then we can write the sum in (3.6) as an iterated sum where the outer sum is over the equivalence classes. 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)

$$\mathcal{W}_h = \{w_{-d}^k, \dots, w_0^k, \dots, w_d^k: k = 1, \dots, h\}.$$

Any minimal d matched $\mathbf{a} = (a_1, \dots, a_{2h})$ induces a partition as given in (3.5). With this \mathbf{a} , associate the *word* $w = w[1]w[2] \cdots w[2h]$ of length $2h$ where

$$w[i_k] = w_0^k, \quad w[j_k] = w_l^k \quad \text{if } a_{i_k} - a_{j_k} = l, 1 \leq k \leq h. \tag{3.7}$$

As an example, consider $d = 1, h = 3$ and $\mathbf{a} = (a_1, \dots, a_6) = (1, 21, 1, 20, 39, 40)$. Then the unique partition of $\{1, 2, \dots, 6\}$ and the unique word associated with \mathbf{a} are $\{\{1, 3\}, \{2, 4\}, \{5, 6\}\}$ and $[w_0^1 w_0^2 w_0^1 w_1^2 w_0^3 w_{-1}^3]$, respectively.

Note that corresponding to any fixed partition $\mathcal{P} = \{\{i_k, j_k\}, 1 \leq k \leq h\}$, 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 $w_0^1 w_i^1 w_0^2 w_j^2$ 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 $w[y]$ (say $w[x] \approx w[y]$) 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 $|a_x - a_y| \leq d$).

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 such a w , the corresponding equivalence class is given by

$$\begin{aligned} \Pi(w) &= \{(\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{A}: w[i_k] = w_0^k, w[j_k] = w_l^k \\ &\Leftrightarrow \mathbf{a}(\mathbf{t}, \boldsymbol{\pi})_{i_k} - a(\mathbf{t}, \boldsymbol{\pi})_{j_k} = l \text{ and } \mathbb{I}_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} = 1\}. \end{aligned} \tag{3.8}$$

Then we rewrite (3.6) as (provided the second limit exists)

$$\lim_{n \rightarrow \infty} \mathbb{E}[\beta_h(\Gamma_{n,d})] = \sum_{\mathcal{P}} \sum_{w \in \mathcal{P}} \lim_{n \rightarrow \infty} \frac{1}{n^{h+1}} \sum_{(t, \pi) \in \Pi(w)} \mathbb{E}[X_{\mathbf{a}(t, \pi)} \mathbb{I}_{\mathbf{a}(t, \pi)}]. \tag{3.9}$$

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

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

Using the definitions of $\gamma_{X^{(d)}}(\cdot)$ and of $S_{h,d}$ given in (2.3), we rewrite (3.9) as (for any set Z , $\#Z$ denotes the number of elements in Z)

$$\lim_{n \rightarrow \infty} \mathbb{E}[\beta_h(\Gamma_{n,d})] = \sum_{\mathcal{P}} \sum_{S_{h,d}} \sum_{w \in \mathcal{P} \cap \mathcal{W}(\mathbf{k})} \lim_{n \rightarrow \infty} \frac{1}{n^{h+1}} \#\Pi(w) \prod_{i=0}^d [\gamma_{X^{(d)}}(i)]^{k_i} \tag{3.10}$$

provided the following limit exists for every word w of length $2h$.

$$p_w^{(d)} \equiv \lim_{n \rightarrow \infty} \frac{1}{n^{h+1}} \#\Pi(w). \tag{3.11}$$

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

$$\begin{aligned} \Pi^*(w) &= \{(\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{A}: w[i_k] = w_0^k, w[j_k] = w_l^k \\ &\Rightarrow a(\mathbf{t}, \boldsymbol{\pi})_{i_k} - a(\mathbf{t}, \boldsymbol{\pi})_{j_k} = l \text{ and } \mathbb{I}_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} = 1\}. \end{aligned} \tag{3.12}$$

By Lemma 3, we have for every w , $n^{-(h+1)}\#\Pi^*(w) - \Pi(w) \rightarrow 0$. Thus, it is enough to show that $\lim_{n \rightarrow \infty} \frac{1}{n^{h+1}}\#\Pi^*(w)$ exists.

For a pair matched w , we divide its coordinates according to the position of the matches as follows. For $1 \leq i < j \leq h$, let the sets S_i be defined as

$$\begin{aligned} S_1(w) &= \{i: w[i] \approx w[j]\}, & S_2(w) &= \{j: w[i] \approx w[j]\}, \\ S_3(w) &= \{i: w[i] \approx w[j+h]\}, & S_4(w) &= \{j: w[i] \approx w[j+h]\}, \\ S_5(w) &= \{i: w[i+h] \approx w[j+h]\}, & S_6(w) &= \{j: w[i+h] \approx w[j+h]\}. \end{aligned}$$

Let E and $G \subset E$ be defined as

$$\begin{aligned} E &= \{t_1, \dots, t_h, \pi_0, \dots, \pi_h\}, \\ G &= \{t_i | i \in S_1(w) \cup S_3(w)\} \cup \{\pi_0\} \cup \{\pi_i | i+h \in S_5(w)\}. \end{aligned}$$

Elements in G are the indices where any matched letter appears for the *first* time and these will be called the *generating vertices*. G has $(h+1)$ elements say u_1^n, \dots, u_{h+1}^n and for simplicity we will write

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

Claim 1. *Each element of E is a linear expression (say λ_i) of the generating vertices that are all to the left of the element.*

Proof. Let the constants in the proposed linear expressions be $\{m_j\}$.

(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

$$\begin{aligned} \text{for all } i \in S_1(w) \cup S_3(w) \quad \lambda_i &\equiv t_i, \\ \lambda_{h+1} &\equiv \pi_0, \end{aligned}$$

and for all

$$i + h \in S_5(w), \quad \lambda_{i+h+1} \equiv \pi_i.$$

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

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

for some n_j such that $|n_j| \leq d$ and define $m_j = n_j$ for $j \in S_2(w)$ and $\lambda_j \equiv \lambda_i$.

(c) Note that for every π we can write

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

Consider the vector $\mathbf{b} = (b_1, b_2, \dots, b_h) \in \{-1, 1\}^h$. It will be a valid choice if we have

$$b_i(\pi_{i-1} - \pi_i) \geq 0 \quad \text{for all } i. \tag{3.13}$$

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

$$t_i = t_j + b_j(\pi_{j-1} - \pi_j) + n_{j+h} \quad \text{for some integer } n_{j+h} \in \{-d, \dots, 0, \dots, d\}. \tag{3.14}$$

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

$$t_i + |\pi_{i-1} - \pi_i| = t_j + |\pi_{j-1} - \pi_j| + n_{j+h} \quad \text{where } n_{j+h} \in \{-d, \dots, 0, \dots, d\}. \tag{3.15}$$

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

$$\pi_j = \lambda_{j+1+h}^{\mathbf{b}} + m_{j+1+h}, \quad j + h \in S_4(w) \cup S_6(w). \tag{3.16}$$

Hence, inductively, π_j as a linear combination $\{\lambda_j^{\mathbf{b}}\}$ of the generating vertices up to an appropriate constant. The superscript \mathbf{b} emphasizes that $\{\lambda_j^{\mathbf{b}}\}$ depends on \mathbf{b} . Further, $\{\lambda_j^{\mathbf{b}}\}$ depends *only* on the vertices present to the left of it. □

Now we are almost ready to write down an expression for the limit. If λ_i were unique for each \mathbf{b} , then we could write $\#\Pi^*(w)$ as a sum of all possible choices of \mathbf{b} and we could tackle the expression for each \mathbf{b} separately. However, λ_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 can choose any \mathbf{b} . We circumvent this problem as

follows: Let

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

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

Suppose $j + h \in \mathcal{T}$. Define

$$L_j(U_n) := b_j(\lambda_{j+h-1}^{\mathbf{b}}(U_n) - \lambda_{j+h}^{\mathbf{b}}(U_n)) + m_{j+h-1} - m_{j+h} \tag{3.17}$$

$$:= \tilde{L}_j(U_n) + m_{j+h-1} - m_{j+h}. \tag{3.18}$$

Then from (3.14) and (3.15) the region given by (3.13) is

$$\{L_j(U_n) \geq 0\} \equiv \{\tilde{L}_j(U_n) + m_{j+h-1} - m_{j+h} \geq 0\}. \tag{3.19}$$

Claim 2. *The above expression is same for all choices of $\{b_j\}$, for $j + h \in \mathcal{T}$.*

Proof. First, we show that if $j + h \in \mathcal{T}$ then we must have

$$t_j = t_j + |\pi_{j-1} - \pi_j| + n_j \quad \text{for some integer } |n_j| \leq d. \tag{3.20}$$

Suppose this is not true. So first assume that $j + h \in S_6(w)$. Then we will have a relation

$$t_i + b_i(\pi_{i-1} - \pi_i) = t_j + b_j(\pi_{j-1} - \pi_j) + n_j \quad \text{where } i + h \in S_5(w). \tag{3.21}$$

Since $\lambda_j^{\mathbf{b}}$ depends only on the vertices present to the left of it, in (3.21), coefficient of π_i would be non-zero and hence we must have $\lambda_{j+h-1}^{\mathbf{b}} - \lambda_{j+h}^{\mathbf{b}} \neq 0$.

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 (3.20). Now a simple calculation shows that for such relations,

$$b_j(\lambda_{j+h-1}^{\mathbf{b}}(U_n) - \lambda_{j+h}^{\mathbf{b}}(U_n)) + m_{j+h-1} - m_{j+h} = -n_j,$$

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

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 $j + h \in \mathcal{T}$ 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 $\mathbf{b} \in \{-1, 1\}^h$ must satisfy (3.13). Keeping the above in view, 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. For ease of writing, we introduce a few more notation:

$$\begin{aligned}
 \mathbb{I}_{m,h}(U_n) &:= \mathbb{I}(\lambda_{2h+1}^{\mathbf{b}}(U_n) + m_{2h+1} = \lambda_{h+1}^{\mathbf{b}}(U_n) + m_{h+1}), \\
 \mathbb{I}_{\lambda^{\mathbf{b}},L}(U_n) &:= \prod_{j=1}^h \mathbb{I}(\lambda_j^{\mathbf{b}}(U_n) + L_j(U_n) \leq n), \\
 \mathbb{I}_{\lambda^{\mathbf{b}},m}(U_n) &:= \prod_{j=1}^{2h} \mathbb{I}(\lambda_j^{\mathbf{b}}(U_n) + m_j \in \mathcal{N}_n) \quad \text{and} \\
 \mathbb{I}_{\mathcal{T}}(U_n) &:= \prod_{1 \leq j \leq h, j \notin \mathcal{T}} \mathbb{I}(L_j(U_n) \geq 0) \times \prod_{j \in \mathcal{T}} \mathbb{I}(n_j \leq 0).
 \end{aligned}
 \tag{3.22}$$

Now we note that,

$$\begin{aligned}
 p_w^{(d)} &:= \lim_n \frac{1}{n^{h+1}} \#\Pi^*(w) \\
 &= \lim_n \frac{1}{n^{h+1}} \sum_{\mathbf{b} \in \mathcal{B}(w)} \sum_{U_n \in \mathcal{N}_n^{h+1}} \mathbb{I}_{m,h}(U_n) \times \mathbb{I}_{\lambda^{\mathbf{b}},m}(U_n) \times \mathbb{I}_{\lambda^{\mathbf{b}},L}(U_n) \times \mathbb{I}_{\mathcal{T}}(U_n) \\
 &= \lim_n \sum_{\mathbf{b} \in \mathcal{B}(w)} \mathbb{E}_{U_n} [\mathbb{I}_{m,h}(U_n) \times \mathbb{I}_{\lambda^{\mathbf{b}},m}(U_n) \times \mathbb{I}_{\lambda^{\mathbf{b}},L}(U_n) \times \mathbb{I}_{\mathcal{T}}(U_n)].
 \end{aligned}$$

Now it only remains to identify the limit. To this end, first 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, \dots, k_d , the linear functions λ_j 's continue to remain same. The only possible changes will be in the values of m_j 's.

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

Claim 3. *Suppose w is such that $\mathcal{R} := \{\lambda_{2h+1}^{\mathbf{b}}(U_n) + m_{2h+1} = \lambda_{h+1}^{\mathbf{b}}(U_n) + m_{h+1}\}$ 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 $m_j = 0, \forall j$. 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 $m_j = 0$. 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

$$\begin{aligned}
 \lambda_{2h+1}^{\mathbf{b}}(U_n) - \lambda_{h+1}^{\mathbf{b}}(U_n) &\equiv 0 \quad (\text{for } d = 0), \\
 \lambda_{2h+1}^{\mathbf{b}}(U_n) - \lambda_{h+1}^{\mathbf{b}}(U_n) &\equiv 0 \quad \text{and} \quad m_{2h+1} - m_{h+1} = 0 \quad (\text{for general } d).
 \end{aligned}$$

Note that the first relation depends only the partition \mathcal{P} but the second relation is determined by the word w . Now $\lambda_j^{\mathbf{b}}$ being linear forms with integer coefficients

$$\lambda_j^{\mathbf{b}}(U_n) + m_j \in \{1, \dots, n\} \iff \lambda_j^{\mathbf{b}}\left(\frac{U_n}{n}\right) + \frac{m_j}{n} \in (0, 1].$$

Define $\mathbb{I}_{m,h}(U)$, $\mathbb{I}_{\lambda^{\mathbf{b}}, \tilde{L}}(U)$, $\mathbb{I}_{\lambda^{\mathbf{b}}}(U)$ and $\tilde{\mathbb{I}}_{\mathcal{T}}(U)$ as in (3.22) with U_n replaced by U , L replaced by \tilde{L} , \mathcal{N}_n replaced by $(0, 1)$, n replaced by 1, and dropping m_j 's in $\mathbb{I}_{\lambda^{\mathbf{b}}, m}$. Noting $\frac{U_n}{n} \xrightarrow{w} U$ following uniform distribution on $[0, 1]^{h+1}$, $\frac{1}{n^{h+1}} \lim \#\Pi^*(w)$ equals

$$p_w^{(d)} = \sum_{\mathbf{b} \in \mathcal{B}(w)} \mathbb{E}_U [\mathbb{I}_{m,h}(U) \times \mathbb{I}_{\lambda^{\mathbf{b}}, \tilde{L}}(U) \times \mathbb{I}_{\lambda^{\mathbf{b}}}(U) \times \tilde{\mathbb{I}}_{\mathcal{T}}(U)]. \tag{3.23}$$

Now the verification of (C1) is complete by observing that (3.10) becomes

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}[\beta_h(\Gamma_{n,d})] &= \sum_{\mathcal{P}} \sum_{\mathbf{k} \in S_{h,d}} p_{\mathbf{k}}^{\mathcal{P},d} \prod_{i=0}^d [\gamma_{X^{(d)}}(i)]^{k_i} \\ &= \sum_{\mathbf{k} \in S_{h,d}} p_{\mathbf{k}}^{(d)} \prod_{i=0}^d [\gamma_{X^{(d)}}(i)]^{k_i}, \end{aligned} \tag{3.24}$$

where

$$p_{\mathbf{k}}^{\mathcal{P},d} = \sum_{w \in \mathcal{P} \cap \mathcal{W}(\mathbf{k})} p_w^{(d)} \quad \text{and} \quad p_{\mathbf{k}}^{(d)} = \sum_{\mathcal{P}} p_{\mathbf{k}}^{\mathcal{P},d}. \tag{3.25}$$

Since there is no explicit expression for the moments of the LSD, we provide in Table 3 the first three moments of the LSD of $\Gamma_n(X)$, when the input sequence is i.i.d. and MA(1). To calculate the moments, we need to find the contributions $p_w^{(d)}$ for words w . The contributions of different relevant words, are provided in Table 1, and in Table 2, for the i.i.d. case. For the MA(1), one can work out the contributions from there.

Table 1. Contributions from words of length 4 for i.i.d. case

Word w	Contribution $p_w^{(0)}$
aabb	2/3
abab	1
abba	0

Table 2. Contributions from words of length 6 for i.i.d. case

Word w	Contribution $p_w^{(0)}$	Word w	Contribution $p_w^{(0)}$
aabc cb	2/3	abbc ac	1/6
aabb cc	1/6	abcabc	1
aabc bc	1/6	abcacb	0
abab cc	1/6	abcbac	0
abac bc	2/3	abcbca	0
abaccb	1/6	abccab	0
abbacc	2/3	abccba	0
abbcca	1/6		

3.1.3. Verification of (C2) and (C3) for Theorem 2.1(a)

Lemma 4. (a) $\mathbb{E}[n^{-1} \text{Tr}(\Gamma_{n,d}^h) - n^{-1} \mathbb{E}[\text{Tr}(\Gamma_{n,d}^h)]]^4 = O(n^{-2})$. Hence $\frac{1}{n} \text{Tr}(\Gamma_{n,d}^h)$ converges to $\beta_{h,d}$ a.s.

(b) $\{\beta_{h,d}\}_{h \geq 0}$ satisfies (C3) and hence defines a unique probability distribution on \mathbb{R} .

Proof. Proof of part (a) uses ideas from Bryc, Dembo and Jiang [10] but the inputs of the matrix are no longer independent, and therefore some modifications are needed. Details are available in Basak, Bose and Sen [5].

(b) Using (3.24) and (2.4) and noting that the number of ways of choosing the partition $\{1, \dots, 2h\} = \bigcup_{i=1}^h \{i_l, j_i\}$ for $\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})$ is $\frac{(2h)!}{2^h h!}$, it easily follows that

$$\begin{aligned}
 |\beta_{h,d}| &\leq \sum_{S_{h,d}} \frac{4^h (2h)!}{h!} \frac{h!}{k_0! \dots k_d!} \prod_{i=0}^d |\gamma_{X^{(d)}}(i)|^{k_i} \\
 &\leq \frac{4^h (2h)!}{h!} \left(\sum_{j=0}^d \sum_{k=0}^{d-j} |\theta_k \theta_{k+j}| \right)^h \leq \frac{4^h (2h)!}{h!} \left(\sum_{k=0}^d |\theta_k| \right)^{2h}.
 \end{aligned}
 \tag{3.26}$$

This implies (C3) holds, proving the lemma. Proof of Theorem 2.1(a) is now complete. □

Table 3. First three moments for i.i.d. and MA(1) input sequence

	i.i.d.	MA(1)
Mean	θ_0^2	$\theta_0^2 + \theta_1^2$
Second moment	$\frac{5}{3} \theta_0^4$	$\frac{5}{3} (\theta_0^2 + \theta_1^2)^2 + \frac{20}{3} \theta_0^2 \theta_1^2$
Third moment	$4\theta_0^6$	$4(\theta_0^2 + \theta_1^2)^3 + 24(\theta_0^2 + \theta_1^2)(2\theta_0\theta_1)^2$

3.1.4. Proof of Theorem 2.1(b) (infinite order case)

First, we assume $\{\varepsilon_t\}$ is i.i.d. Fix $\varepsilon > 0$. Choose d such that $\sum_{k \geq d+1} |\theta_k| \leq \varepsilon$. For convenience we will write $\Gamma_n(X) = \Gamma_n$. Clearly, $\Gamma_n = A_n A_n^T$ where

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

By ergodic theorem, a.s., we have the following two relations:

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

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

Hence using Lemma 1(b), a.s.

$$\limsup_n d_{\text{BL}}^2(F^{\Gamma_{n,d}}, F^{\Gamma_n}) \leq 2 \left(\sum_{k=0}^{\infty} |\theta_k| \right)^2 \varepsilon^2. \tag{3.27}$$

Now $F^{\Gamma_{n,d}} \xrightarrow{w} F_d$ a.s. Since d_{BL} metrizes weak convergence of probability measures as $n \rightarrow \infty$, $d_{\text{BL}}(F^{\Gamma_{n,d}}, F_d) \rightarrow 0$, a.s. Since $\{F^{\Gamma_{n,d}}\}_{n \geq 1}$ is Cauchy with respect to d_{BL} a.s., by triangle inequality, and (3.27), $\limsup_{m,n} d_{\text{BL}}(F^{\Gamma_n}, F^{\Gamma_m}) \leq 2\sqrt{2}(\sum_{k=0}^{\infty} |\theta_k|)\varepsilon$. Hence $\{F^{\Gamma_n}\}_{n \geq 1}$ is Cauchy with respect to d_{BL} a.s. Since d_{BL} is complete, there exists a probability measure F on \mathbb{R} such that $F^{\Gamma_n} \xrightarrow{w} F$ a.s. Further

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

and hence $F_d \xrightarrow{w} F$ as $d \rightarrow \infty$. Since $\{F_d\}$ are non-random, F is also non-random.

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. We omit the details.

To show convergence of $\{\beta_{h,d}\}$, we note that under Assumption B(b), (3.26) yields

$$\sup_d |\beta_{h,d}| \leq c_h := \frac{4^h (2h)!}{h!} \left(\sum_{k=0}^{\infty} |\theta_k| \right)^{2h} < \infty \quad \forall h \geq 0. \tag{3.28}$$

Hence for every fixed h , $\{A_d^h\}$ is uniformly integrable where $A_d \sim F_d$. Since $F_d \xrightarrow{w} F$,

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

completing the proof of (b). Since $|\beta_h| \leq c_h$, it easily follows that $\{\beta_h\}_{h \geq 0}$ satisfies (C3) and hence uniquely determines the distribution F .

3.1.5. Proof of Theorem 2.1(c)

We first claim that for $d \geq 0$ $p_{k_0, \dots, k_d}^{(d)} = p_{k_0, \dots, k_d, 0}^{(d+1)}$. To see this, 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 $\# \{ (\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{A} | \mathbf{a}(\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{M} \} = O(n^h)$. Hence,

$$p_{k_0, \dots, k_d}^{(d)}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n^{h+1}} \# \left\{ (\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{A} | \mathbf{a}(\mathbf{t}, \boldsymbol{\pi}) \text{ is minimal } d \text{ matched}$$

$$\text{with partition } \{1, \dots, 2h\} = \bigcup_{l=1}^h \{i_l, j_l\}$$

$$\text{and there are exactly } k_s \text{ many } l\text{'s for which}$$

$$| \mathbf{a}(\mathbf{t}, \boldsymbol{\pi})(i_l) - \mathbf{a}(\mathbf{t}, \boldsymbol{\pi})(j_l) | = s, s = 0, \dots, d, \mathbb{I}_{\mathbf{a}(\mathbf{t}, \boldsymbol{\pi})} = 1 \text{ and}$$

$$| \mathbf{a}(\mathbf{t}, \boldsymbol{\pi})(x) - \mathbf{a}(\mathbf{t}, \boldsymbol{\pi})(y) | \geq d + 2 \text{ if } x, y \text{ belong to}$$

$$\text{different partition blocks} \right\}$$

$$= p_{k_0, \dots, k_d, 0}^{(d+1)}.$$

Thus for $\theta_0, \dots, \theta_d \geq 0$ and $d \geq 1$,

$$\beta_{h,d} \geq \sum_{S_{h,d-1}} p_{k_0, \dots, k_{d-1}, 0}^{(d)} \prod_{i=0}^{d-1} [\gamma_{X^{(d)}}(i)]^{k_i}$$

$$\geq \sum_{S_{h,d-1}} p_{k_0, \dots, k_{d-1}}^{(d-1)} \prod_{i=0}^{d-1} [\gamma_{X^{(d-1)}}(i)]^{k_i} = \beta_{h,d-1},$$

proving the result.

Incidentally, if Assumption B(a) is violated, then the ordering need not hold. This can be checked by considering an MA(2) and an MA(1) process with parameters $\theta_0, \theta_1, \theta_2$ and where $\theta_2 = -\kappa\theta_0, \theta_0, \theta_1 > 0$. Then $\beta_{2,2} < \beta_{2,1}$ if we choose $\kappa > 0$ sufficiently small. The details are available in Basak, Bose and Sen [5].

3.1.6. Proof of unbounded support of F_d and F

For any word w , let $|w|$ denote the length of the word. Let

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

Then

$$\beta_{2h,d} \geq [\gamma_{X^{(d)}}(0)]^{2h} p_{2h,0,\dots,0} \geq [\gamma_{X^{(d)}}(0)]^{2h} \sum_{w \in \mathcal{W}} \lim_n n^{-(2h+1)} \#\Pi^*(w). \tag{3.29}$$

For $w = w_1 w_2 \in \mathcal{W}$, let $\{1, \dots, 2h\} = \bigcup_{i=1}^h (i_s, j_s)$ be the partition corresponding to w_1 . Then

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

Now adapting the ideas of Bryc, Dembo and Jiang [10], we obtain that for each d finite F_d has unbounded support. Since $\{\beta_{h,d}\}$ increases to β_h , same conclusion is true for F . For details see Basak, Bose and Sen [5].

3.2. Outline of the proof of Theorem 2.3

3.2.1. Proof of Theorem 2.3(a), (b) 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 ACVMs with parameter α . We begin by noting that the expression for these contain an extra indicator term $\mathbb{I}_1 = \prod_{i=1}^h \mathbb{I}(|\pi_{i-1} - \pi_i| \leq m_n)$ and $\mathbb{I}_2 = \prod_{i=1}^h \mathbb{I}(1 \leq \pi_i \leq m_n)$, respectively. For type II ACVMs since there are m_n eigenvalues instead of n , the normalising denominator is now m_n . Hence,

$$\beta_h(\Gamma_{n,d}^{\alpha,I}) = \frac{1}{n^{h+1}} \sum_{\substack{1 \leq \pi_0, \dots, \pi_h \leq n \\ \pi_h = \pi_0}} \left[\prod_{j=1}^h \left(\sum_{t_j=1}^n X_{t_j,d} X_{t_j+|\pi_j-\pi_{j-1}|,d} \mathbb{I}(t_j+|\pi_j-\pi_{j-1}| \leq n) \right) \right] \mathbb{I}_1$$

and

$$\frac{m_n}{n} \beta_h(\Gamma_{n,d}^{\alpha,II}) = \frac{1}{n^{h+1}} \sum_{\substack{1 \leq \pi_0, \dots, \pi_h \leq n \\ \pi_h = \pi_0}} \left[\prod_{j=1}^h \left(\sum_{t_j=1}^n X_{t_j,d} X_{t_j+|\pi_j-\pi_{j-1}|,d} \mathbb{I}(t_j+|\pi_j-\pi_{j-1}| \leq n) \right) \right] \mathbb{I}_2.$$

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 2.1.

Since there are only the extra indicator terms, the negligibility of higher order edges and verification of (C2) and (C3) needs no new arguments. Likewise, verification of (C1) is also similar except that there is now an extra indicator term in the expression for $p_w^{(d)}$. This takes care of the finite d case. For $d = \infty$, note that the type II ACVMs are $m_n \times m_n$ principal subminor of the original sample ACVMs and hence are automatically non-negative definite. 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 imitating the proof of Theorem 2.1, we can move from finite d to $d = \infty$. However for type I ACVMs, we cannot apply these arguments, as these matrices are not necessarily non-negative definite. Rather we proceed as in the proof of Theorem 2.2. Previous proof of unbounded support now needs only minor changes. We omit the details.

Since $\Gamma_{n,d}^{\alpha, II}$ is non-negative definite, the technique of proof of Theorem 2.1 can be adopted under Assumption A(a).

3.2.2. Proof of Theorem 2.3(b) for type I band ACVM

Existence: Let $p_w^{(d),0,I}$ be the limiting contribution of the word w for type I ACVM with band parameter $\alpha = 0$. Then

$$p_w^{(d),0,I} := \lim_n \frac{1}{n^{h+1}} \sum_{\mathbf{b} \in \mathcal{B}(w)} \mathbb{E}_{U_n} [\mathbb{I}_{m,h}(U_n) \times \mathbb{I}_{\lambda^{\mathbf{b}},m}(U_n) \times \mathbb{I}_{\lambda^{\mathbf{b}},L}(U_n) \times \mathbb{I}_{\mathcal{T}}^I(U_n)],$$

where

$$\mathbb{I}_{\mathcal{T}}^I(U_n) = \mathbb{I}_{\mathcal{T},L}(U_n) \times \mathbb{I}_{\mathcal{T},m} := \prod_{\substack{j=1 \\ j \notin \mathcal{T}}}^h \mathbb{I}(0 \leq L_j(U_n) \leq m_n) \times \prod_{j \in \mathcal{T}} \mathbb{I}(-m_n \leq n_j \leq 0).$$

If w , $\lambda_{j+h-1}^{\mathbf{b}} \neq \lambda_{j+h}^{\mathbf{b}}$ for some j , then $\mathbb{I}_{\mathcal{T},L}(U_n) \rightarrow 0$ as $n \rightarrow \infty$ and thus limiting contribution from that word will be 0. Thus, only those words w for which $\lambda_{h+1}^{\mathbf{b}} = \lambda_{j+h}^{\mathbf{b}}$ for all $j \in \{1, 2, \dots, h+1\}$ may contribute non-zero 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 of Basak, Bose and Sen [5], and the expression for limiting moments for $\Gamma_n(X)$ shows that $w \in \mathcal{W}_0^h$ may contribute non-zero quantity, where

$$\mathcal{W}_0^h = \{w: |w| = 2h, w[i] \text{ matches with } w[i+h], n_i \leq 0, i = 1, 2, \dots, h\}.$$

Further note that if $w \in \mathcal{W}_0^h$ then $\mathcal{T} = \{h+1, h+2, \dots, 2h\}$, and thus $\mathbb{I}_{\mathcal{T},L} \equiv 1$.

For $d = 0$ note that $\#\mathcal{W}_0^h = 1$ for every h and one can easily check that the contribution from that word is 1. Thus $\beta_{h,0}^0 = \theta_0^{2h}$ 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, and if $m_n \geq d$, then

$$\mathbb{I}_{\lambda^{\mathbf{b}},m} \times \mathbb{I}_{\lambda^{\mathbf{b}},L} \times \mathbb{I}_{\mathcal{T},m} \rightarrow \prod_{j=1}^h \mathbb{I}(n_j \leq 0) \quad \text{as } n \rightarrow \infty.$$

Combining the above arguments we get that for any $w \in \mathcal{W}_0^h$, $p_w^{(d),0,I}$ is the number of choices of $\mathbf{b} \in \mathcal{B}(w)$, and $\{n_1, n_2, \dots, n_h; n_i \leq 0\}$, such that $\sum_i n_i b_i = 0$.

Noting that type I ACVMs are not necessarily non-negative definite, we need to adapt the proof of Theorem 2.2. Details are omitted.

Identification of the LSD: Now it remains to argue that the limit we obtained is same as $f_X(U)$. For $d = 0$ LSD is $\delta_{\theta_0^2}$ and it is trivial to check it is same as $f_X(U)$.

For $0 < d < \infty$, note that the proof does not use the fact that $m_n \rightarrow \infty$ and we further note that for any sequence $\{m_n\}$ the limit we obtained above will be same whenever $\liminf_{n \rightarrow \infty} m_n \geq d$. So in particular the limit will be same if we choose another sequence $\{m'_n\}$ such that $m'_n = d$ for all n . Let $\Gamma_{n',d}^I$ denote the type I ACVM 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

$$\begin{aligned} d_{\text{BL}}^2(F^{\Gamma_{n',d}^I}, F^{\Sigma_{n,d}}) &\leq \frac{1}{n} \text{Tr}(\Gamma_{n',d}^I - \Sigma_{n,d})^2 \\ &\leq 2(\hat{\gamma}_{X^{(d)}}(0) - \gamma_{X^{(d)}}(0))^2 + \dots + 2(\hat{\gamma}_{X^{(d)}}(d) - \gamma_{X^{(d)}}(d))^2. \end{aligned}$$

For any j as $n \rightarrow \infty$, $\hat{\gamma}_{X^{(d)}}(j) \rightarrow \gamma_{X^{(d)}}(j)$ a.s. Since d is finite, the right side of the above expression goes to 0 a.s. This proves the claim for d finite.

To prove the result for the case $d = \infty$, first note that we already have

$$\text{LSD}(\Gamma_{n,d}^{0,I}) = \text{LSD}(\Sigma_{n,d}) := G_d \quad \text{and} \quad \text{LSD}(\Gamma_{n,d}^{0,I}) \xrightarrow{w} \text{LSD}(\Gamma_n^{0,I}) \quad \text{as } d \rightarrow \infty.$$

Thus, it is enough to prove that $G_d \xrightarrow{w} G (= \text{LSD}(\Sigma_n))$ as $d \rightarrow \infty$ where Σ_n is the $n \times n$ matrix whose (i, j) th entry is $\gamma_X(|i - j|)$. Define a 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. By triangle inequality,

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

Fix any $\varepsilon > 0$. Fix d_0 such that $(\sum_{j=0}^{\infty} |\theta_j|)^2 (\sum_{l=d+1}^{\infty} |\theta_l|)^2 \leq \frac{\varepsilon^2}{32}$ for all $d \geq d_0$. Now again using Lemma 1(a) we get the following two relations:

$$\begin{aligned} \limsup_n d_{\text{BL}}^2(F^{\Sigma_{n,d}}, F^{\bar{\Sigma}_{n,d}}) &\leq 2[(\gamma_{X^{(d)}}(0) - \gamma_X(0))^2 + \dots + (\gamma_{X^{(d)}}(d) - \gamma_X(d))^2] \\ &= 2 \sum_{j=0}^d \left(\sum_{k=d-j+1}^{\infty} \theta_k \theta_{j+k} \right)^2 \leq \frac{\varepsilon^2}{16}, \\ d_{\text{BL}}^2(F^{\bar{\Sigma}_{n,d}}, F^{\Sigma_n}) &\leq \limsup_n \frac{1}{n} \text{Tr}(\bar{\Sigma}_{n,d} - \Sigma_n)^2 \leq \frac{\varepsilon^2}{16}. \end{aligned}$$

Thus, $\limsup_n d_{\text{BL}}(F^{\Sigma_{n,d}}, F^{\Sigma_n}) \leq \varepsilon/2$, for any $d \geq d_0$, and therefore by triangle inequality, $d_{\text{BL}}(F^{G_d}, F^G) \leq \varepsilon$. This completes the proof.

3.2.3. Proof of Theorem 2.3(b) for type II band autocovariance matrix

First, note that by Lemma 3 we need to consider only minimal matched terms. Let

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

Since $1 \leq \pi_i \leq m_n$ for all i , by similar arguments as in Lemma 3 we get

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

Thus, for any word w such that $\#G_t < h$ the limiting contribution will be 0. Hence only contributing words e in this case are those for which $\#S_3(w) = \#S_4(w) = h$. and from Lemma 6 of Basak, Bose and Sen [5], the only contributing words are those belonging to \mathcal{W}_0^h . Therefore using same arguments as in the proof of Theorem 2.3, for type I ACVM, for $\alpha = 0$ we obtain the same limit. All the remaining conclusions here follow from the proof for type I ACVMs with parameter $\alpha = 0$.

Since type II ACVMs are non-negative definite, connection between the LSD for finite d and $d = \infty$ is proved adapting the ideas from the proof of Theorem 2.1.

3.2.4. Proof of Theorem 2.3(c)

Since K is bounded, negligibility of higher order edges and verification of (C2) and (C3) is same as before. Verification of (C1) is also same, with an extra indicator in the limiting expression. Denoting $p_w^{(d),K}$ to be the limiting contribution from a word w , we have,

$$p_w^{(d),K} = \lim_n \mathbb{E}_{U_n} [\mathbb{I}_{m,h}(U_n) \times \mathbb{I}_{\lambda^b,m}(U_n) \times \mathbb{I}_{\lambda^b,L}(U_n) \times \mathbb{I}_{\mathcal{T}}^{\mathbb{I}}(U_n) \times \mathbb{I}_K(U_n)],$$

where

$$\mathbb{I}_K(U_n) := \prod_{j=1}^h K\left(\frac{L_j(U_n)}{m_n}\right).$$

Since $m_n \rightarrow \infty$, and $K(\cdot)$ is continuous at 0, $K(0) = 1$, note that $\mathbb{I}_K \rightarrow 1$. Now arguing as in Section 3.2.2, 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 are similar as in Section 3.2.2 and the details are omitted.

3.3. Proof of Theorem 2.2

Proceeding as earlier it is easy to see the limit exists, and for each word w , the limiting contribution is given by,

$$p_w^{*,(d)} = \sum_{\mathbf{b} \in \mathcal{B}(w)} \mathbb{E}_U [\mathbb{I}_{m,h}(U) \times \mathbb{I}_{\lambda^b}(U) \times \tilde{I}_{\mathcal{T}}(U)].$$

Comparing the above expression with the corresponding expression for the sequence $\Gamma_{n,d}$,

$$\beta_{h,d} \leq \beta_{h,d}^* \quad \text{if } \theta_j \geq 0, 0 \leq j \leq d.$$

Relation (3.26) 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 thus also Carleman's condition is satisfied.

Since Γ_n^* and $\Gamma_{n,d}^*$ are no longer positive definite matrices the ideas used in the proof of Theorem 2.1(b) cannot be adapted here. We proceed as follows instead: Note that

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

Write

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

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

$$\mathbb{E}[\beta_h(\Gamma_n^*)] = \sum_{\substack{k_j, k'_j \geq 0 \\ j=1, \dots, h}} \prod_{j=1}^h (\theta_{k_j} \theta_{k'_j}) \frac{1}{n^{h+1}} \mathbb{E} \left[\sum_{(\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{A}} \prod_{j=1}^h \varepsilon_{t_j - k_j} \varepsilon_{t_j + |\pi_j - \pi_{j-1}| - k'_j} \right].$$

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.

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

One can proceed as in the proof of Theorem 2.1 to show that only pair matched words contribute and hence enough to argue that $\lim_n n^{-(h+1)} \#\{(\mathbf{t}, \boldsymbol{\pi}) \in \mathcal{A} : \{t_j - k_j, t_j + |\pi_j - \pi_{j-1}| - k'_j, j = 1, \dots, h\} \text{ is pair matched}\}$ exists, and which follows by adapting the ideas used in the proof of Theorem 2.1. Note that appropriate compatibility is needed among $\{k_j, k'_j, j = 1, \dots, h\}$, 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 depend on $\{k_j, k'_j, j = 1, \dots, h\}$.

We also note that

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

Hence, F^* is uniquely determined by its moments and using DCT, $\beta_{h,d}^* \rightarrow \beta_h^*$. Whence it also follows that $F_d^* \xrightarrow{w} F^*$. Proof of part (c) is similar to the proof of Theorem 2.1(c). \square

Remark 3.1. Theorem 2.2 has not been proved under Assumption A(a) because there is no straightforward way to apply (3.1) or (3.2) since $\Gamma_n^*(X)$ is not non-negative definite. Simulation results indicate that the same LSD continues to hold under Assumption A(a).

Acknowledgements

We thank Dimitris Politis and Mohsen Pourahmadi for sharing their work and thoughts. The constructive comments of the four Referees and the Associate Editor is gratefully acknowledged. We thank the Editor for his encouragement to submit a revision.

A. Basak supported by Melvin and Joan Lane endowed Stanford Graduate Fellowship fund. A. Bose's research supported by J.C. Bose Fellowship, Govt. of India. S. Sen supported by NYU graduate fellowship under Henry M. MacCracken Program.

Supplementary Material

Simulations (DOI: [10.3150/13-BEJ520SUPP](https://doi.org/10.3150/13-BEJ520SUPP); .pdf). Recall that none of the LSDs have a nice description. Following the suggestion of one of the Referees, we have collected some simulation results in a supplementary file Basak, Bose and Sen [6].

The simulations are for the AR(1) and MA(1) models. These simulations provide evidence that the limits are indeed universal and exhibit some mass on the negative axis for the ESD (and hence the LSD) of $\Gamma_n^*(X)$. They also show how the LSD of type I banded $\Gamma_n(X)$ changes with the model as well as the value of the parameter α . The unbounded nature of the LSD is also evident from these simulations.

For the banded matrices, the simulations demonstrate that for small values of α , the LSD of $\Sigma_n(X)$ and $\Gamma_n(X)$ are virtually indistinguishable for large n , confirming that thinly banded ACVMs are consistent for $\Sigma_n(X)$. As the value of α increases, the right tail of the LSD thickens, and the probability of being near zero decreases. In general, there may be considerable amount of mass in the negative axis. This mass reduces as the value of α decreases.

The LSD of $\Gamma_n(X)$ varies as the parameter of the models change. For both AR(1) and MA(1) models, as θ increases from 0, the tail thickens, and the mass near zero decreases. For the AR(1) model, when θ approaches 1, that is, when the process is near non-stationary the LSD becomes very flat, and its tail becomes huge.

References

- [1] Arcones, M.A. (2000). Distributional limit theorems over a stationary Gaussian sequence of random vectors. *Stochastic Process. Appl.* **88** 135–159. [MR1761993](#)
- [2] Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. [MR2411613](#)
- [3] Bai, Z.D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review (with discussions). *Statist. Sinica* **9** 611–677. [MR1711663](#)

- [4] Basak, A. (2009). Large dimensional random matrices. M. Stat. Project report, May 2009. Indian Statistical Institute.
- [5] Basak, A., Bose, A. and Sen, S. (2011). Limiting spectral distribution of sample autocovariance matrices. Technical Report R11 2011. Stat-Math Unit, Indian Statistical Institute. Available at <http://arxiv.org/pdf/1108.3147v1.pdf>.
- [6] Basak, A., Bose, A. and Sen, S. (2013). Supplement to “Limiting spectral distribution of sample autocovariance matrices.” DOI:10.3150/13-BEJ520SUPP.
- [7] Bose, A., Gangopadhyay, S. and Sen, A. (2010). Limiting spectral distribution of XX' matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 677–707. [MR2682263](#)
- [8] Bose, A. and Sen, A. (2008). Another look at the moment method for large dimensional random matrices. *Electron. J. Probab.* **13** 588–628. [MR2399292](#)
- [9] Böttcher, A. and Silbermann, B. (1999). *Introduction to Large Truncated Toeplitz Matrices*. Universitext. New York: Springer. [MR1724795](#)
- [10] Bryc, W., Dembo, A. and Jiang, T. (2006). Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.* **34** 1–38. [MR2206341](#)
- [11] Dudley, R.M. (2002). *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics **74**. Cambridge: Cambridge Univ. Press. Revised reprint of the 1989 original. [MR1932358](#)
- [12] Giraitis, L., Robinson, P.M. and Surgailis, D. (2000). A model for long memory conditional heteroscedasticity. *Ann. Appl. Probab.* **10** 1002–1024. [MR1789986](#)
- [13] Hammond, C. and Miller, S.J. (2005). Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices. *J. Theoret. Probab.* **18** 537–566. [MR2167641](#)
- [14] McMurry, T.L. and Politis, D.N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Series Anal.* **31** 471–482. [MR2732601](#)
- [15] Sen, A. (2006). Large dimensional random matrices. M. Stat. Project report, May 2006. Indian Statistical Institute.
- [16] Sen, S. (2010). Limiting spectral distribution of random matrices. M. Stat. Project report, July 2010. Indian Statistical Institute.
- [17] Wu, W.B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768. [MR2589209](#)
- [18] Xiao, H. and Wu, W.B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40** 466–493. [MR3014314](#)
- [19] Yao, J. (2012). A note on a Marčenko–Pastur type theorem for time series. *Statist. Probab. Lett.* **82** 22–28. [MR2863018](#)

Received September 2011 and revised December 2012