

Research Article
Special Issue: Combinatorial Matrices

Open Access

Marianna Bolla*, Tamás Szabados, Máté Baranyi, and Fatma Abdelkhalek

Block circulant matrices and the spectra of multivariate stationary sequences

<https://doi.org/10.1515/spma-2020-0121>

Received October 30, 2020; accepted December 16, 2020

Abstract: Given a weakly stationary, multivariate time series with absolutely summable autocovariances, asymptotic relation is proved between the eigenvalues of the block Toeplitz matrix of the first n autocovariances and the union of spectra of the spectral density matrices at the n Fourier frequencies, as $n \rightarrow \infty$. For the proof, eigenvalues and eigenvectors of block circulant matrices are used. The proved theorem has important consequences as for the analogies between the time and frequency domain calculations. In particular, the complex principal components are used for low-rank approximation of the process; whereas, the block Cholesky decomposition of the block Toeplitz matrix gives rise to dimension reduction within the innovation subspaces. The results are illustrated on a financial time series.

Keywords: weakly stationary time series, autocovariance and spectral density matrix, finite Fourier transform, complex principal components, block Cholesky decomposition

MSC: 42B05, 62M15, 65F30

1 Introduction

Let $\{X_t\}$ be a weakly stationary, real-valued time series ($t \in \mathbb{Z}$). Assume that $\mathbb{E}(X_t) = 0$, $c(0) = \mathbb{E}(X_t^2) > 0$, and the sequence of autocovariances $c(h)$, ($h = 1, 2, \dots$) is absolutely summable; obviously, $c(h) = c(-h)$. Then it is known that the Toeplitz matrix $\mathbf{C}_n = [c(i-j)]_{i,j=1}^n$, that is the covariance matrix of the random vector $(X_1, \dots, X_n)^T$, is positive definite for all $n \in \mathbb{N}$ (the vectors are column vectors and T denotes the transposition). Under the absolute summability condition, the spectral density function f of the process exists and it is continuous. We can as well use $[-\pi, \pi]$ or $[0, 2\pi]$ for a complete period of f , and in view of $f(\omega) = f(-\omega) = f(2\pi - \omega)$, it suffices to confine ourselves to the $[0, \pi]$ interval. In [1], an asymptotic relation between the eigenvalues of \mathbf{C}_n and the values $f(\omega_j)$ is proved as $n \rightarrow \infty$, where $\omega_j = j \frac{2\pi}{n}$ is the j th Fourier frequency, $j = 0, 1, \dots, n-1$.

In the present paper we prove a similar, albeit more complicated, result in the multivariate case. Let $\{\mathbf{X}_t\}$ be a weakly stationary d -dimensional time series, $t \in \mathbb{Z}$ and the state space is \mathbb{R}^d . Again, $\mathbb{E}(\mathbf{X}_t) = \mathbf{0}$ and the sequence of the $d \times d$ autocovariance matrices $\mathbf{C}(0), \mathbf{C}(1), \mathbf{C}(2), \dots$ is absolutely summable (entrywise); also $\mathbf{C}(-h) = \mathbf{C}^T(h)$ and $\mathbf{C}(0)$ is the usual covariance matrix of \mathbf{X}_t . Now the $nd \times nd$ covariance matrix \mathbf{C}_n of the compounded random vector $(\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)^T$ is the following symmetric, positive semidefinite block-Toeplitz

***Corresponding Author: Marianna Bolla:** Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics Muegyetem rkp. 3. Bldg H/5. H-1111 Budapest, Hungary, E-mail: marib@math.bme.hu

Tamás Szabados: Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics Muegyetem rkp. 3. Bldg H/5. H-1111 Budapest, Hungary, E-mail: szabados@math.bme.hu

Máté Baranyi: Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics Muegyetem rkp. 3. Bldg H/5. H-1111 Budapest, Hungary, E-mail: baranyim@math.bme.hu

Fatma Abdelkhalek: Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics Muegyetem rkp. 3. Bldg H/5. H-1111 Budapest, Hungary, E-mail: fatma@math.bme.hu

matrix:

$$\mathfrak{C}_n := \begin{bmatrix} \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \cdots & \mathbf{C}(n-1) \\ \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \cdots & \mathbf{C}(n-2) \\ \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \cdots & \mathbf{C}(n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}^T(n-1) & \mathbf{C}^T(n-2) & \mathbf{C}^T(n-3) & \cdots & \mathbf{C}(0) \end{bmatrix}, \quad (1)$$

the (i, j) block of which is $\mathbf{C}(j-i)$. The symmetry comes from the fact, that the (j, i) block is $\mathbf{C}(i-j) = \mathbf{C}^T(j-i)$.

At the same time, we consider the (under our conditions, existing) spectral density matrix \mathbf{f} of the process. This is a self-adjoint, positive semidefinite, $d \times d$ matrix, which is, in fact, the Fourier transform of the autocovariance matrices:

$$\mathbf{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbf{C}(h)e^{-ih\omega}, \quad \omega \in [0, 2\pi].$$

It has real diagonal, but usually complex off-diagonal entries. In view of $\mathbf{f}(-\omega) = \overline{\mathbf{f}(\omega)}$, it suffices to confine ourselves to the $[0, \pi]$ interval (here $\bar{\cdot}$ denotes the complex, entrywise conjugation).

In our main theorem, we state that for ‘large’ n , the eigenvalues of \mathfrak{C}_n asymptotically comprise the union of the spectra of the spectral density matrices $\mathbf{f}(\omega_j)$, $j = 0, 1, \dots, n-1$. (We use spectra in two different meanings here: for the eigenvalues and for the spectral representation of weakly stationary processes.) The theorem also has the computational benefit that instead of finding the eigenvalues of an $nd \times nd$ matrix (that needs $\mathcal{O}(n^3 d^3)$ operations) we can find the eigenvalues of number n of $d \times d$ matrices (that means only $\mathcal{O}(nd^3)$ operations).

Under our conditions, the rank r of $\mathbf{f}(\omega)$ is almost surely constant on $[0, \pi]$, but it can be less than d . It is true for sliding summations (two-sided moving average processes), in particular, for regular ones (one-sided moving average processes). Processes of rational spectral density, e.g., VARMA processes, are such. At the same time, the dimension of the innovation subspaces, which equals to r , can be concluded from the diagonal blocks of the block Cholesky decomposition of \mathfrak{C}_n , and it can also be reduced in the time domain with usual principal component (PC) analysis techniques. In parallel, in the frequency domain, an m -rank approximation with $m \leq r$ is also possible by using PC transformation with m complex PCs, where m depends on the gap of the spectrum of \mathbf{f} , see also [2].

The organization of the paper is as follows. In Section 2, we introduce block circulant matrices and show how existing results for their spectra are applicable in our situation. It is important, that because of the duplication of the eigenvalues, eigenvectors of both real and complex coordinates are available. Both kinds of eigenvectors are intensively studied. After this preparation, we prove our main theorem in Section 3. Section 4 discusses consequences of this theorem, like bounds for the spectra, complex PC transformation, low rank approximation of the process, illustrated on real-life financial data. Eventually, in Section 5, conclusions are drawn.

2 Preliminaries

To characterize the eigenvalues of the block-Toeplitz matrix \mathfrak{C}_n , we need the symmetric block circulant matrix $\mathfrak{C}_n^{(s)}$ that we consider for odd n , say $n = 2k + 1$ here (for even n , the calculations are similar); for the definition, see [3]. In fact, the rows of a circulant matrix are cyclic permutations of the preceding ones; whereas, in the block circulant case, when permuting, we take transposes of the blocks if those are not symmetric themselves, see the example of Equation (2). Spectra of block circulant matrices are well characterized, but \mathfrak{C}_n is not block circulant, in general; this is why $\mathfrak{C}_n^{(s)}$ is constructed, by disregarding the autocovariances of order greater than $\frac{n}{2}$. This can be done only on the assumption that the sequences of autocovariances are (entrywise) absolutely summable. It is crucial that k is the integer part of $\frac{n}{2}$, as the constructed block circulant matrix is symmetric only with this choice of k . (In case of even n , the choice $\frac{n}{2}$ is perfect, and it causes only

minor changes in the spectrum, as will be discussed later.) This choice also has benefits when the autocovariance matrices are estimated based on an n -length long time series $\mathbf{X}_1, \dots, \mathbf{X}_n$:

$$\hat{\mathbf{C}}(h) = \hat{\mathbf{C}}^*(-h) = \sum_{t=1}^{n-h} \mathbf{X}_{t+h} \mathbf{X}_t^*, \quad h = 0, 1, \dots, \lfloor \frac{n}{2} \rfloor$$

if $\mathbb{E}\mathbf{X}_t = \mathbf{0}$, otherwise the estimated mean is subtracted. Here at least $\frac{n}{2}$ sample entries are used for each $\hat{\mathbf{C}}(h)$ with $h \leq \frac{n}{2}$ that makes the estimate more accurate. (Note that the above estimate is asymptotically unbiased even though its ergodicity is guaranteed only if $h = o(n)$ as $n \rightarrow \infty$. Though $\frac{n}{2}$ is obviously not $o(n)$, as the proof of the main theorem will show, disregarding the last $\frac{n}{2}$ autocovariances will guarantee the convergence of the spectrum.)

The (i, j) block of $\mathfrak{C}_n^{(s)}$ for $1 \leq i \leq j \leq n$ is

$$\mathfrak{C}_n^{(s)}(\text{block}_i, \text{block}_j) = \begin{cases} \mathbf{C}(j-i), & j-i \leq k \\ \mathbf{C}(n-(j-i)), & j-i > k; \end{cases}$$

whereas, for $i > j$, it is

$$\mathfrak{C}_n^{(s)}(\text{block}_i, \text{block}_j) = \begin{cases} \mathbf{C}^T(i-j), & i-j \leq k \\ \mathbf{C}^T(n-(i-j)), & i-j > k. \end{cases}$$

In this way, $\mathfrak{C}_n^{(s)}$ is a symmetric block Toeplitz matrix, like \mathfrak{C}_n , and it is the same as \mathfrak{C}_n within the blocks (i, j) for which $|j-i| \leq k$ holds. However, $\mathfrak{C}_n^{(s)}$ is also a block circulant matrix that fits our purposes. For example, if $n = 7$ and $k = 3$, then we have

$$\mathfrak{C}_7^{(s)} := \begin{bmatrix} \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \mathbf{C}(3) & \mathbf{C}(3) & \mathbf{C}(2) & \mathbf{C}(1) \\ \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \mathbf{C}(3) & \mathbf{C}(3) & \mathbf{C}(2) \\ \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \mathbf{C}(3) & \mathbf{C}(3) \\ \mathbf{C}^T(3) & \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \mathbf{C}(3) \\ \mathbf{C}^T(3) & \mathbf{C}^T(3) & \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) \\ \mathbf{C}^T(2) & \mathbf{C}^T(3) & \mathbf{C}^T(3) & \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) \\ \mathbf{C}^T(1) & \mathbf{C}^T(2) & \mathbf{C}^T(3) & \mathbf{C}^T(3) & \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) \end{bmatrix}. \quad (2)$$

In the univariate ($d = 1$) case, when $n = 2k + 1$, by Kronecker products (with permutation matrices) it is well known, see e.g., [1, 3, 4], that the j th (real) eigenvalue of $\mathfrak{C}_n^{(s)}$ is

$$\sum_{h=-k}^k c(h) \rho_j^h = c(0) + 2 \sum_{h=1}^k c(h) \cos(h\omega_j),$$

where $\rho_j = e^{i\omega_j}$ is the j th primitive (complex) n th root of 1 and $\omega_j = \frac{2\pi j}{n}$ is the j th Fourier frequency ($j = 0, 1, \dots, n-1$). Further, the eigenvector corresponding to the j th eigenvalue is $(1, \rho_j, \dots, \rho_j^{n-1})^T$; it has norm \sqrt{n} . After normalizing with $\frac{1}{\sqrt{n}}$, we get a complete orthonormal set of eigenvectors (of complex coordinates).

When $\mathbf{C}(h)$ s are $d \times d$ matrices, by inflation techniques and applying Kronecker products, we use blocks instead of entries and the eigenvectors also follow a block structure. In [3, 4], the eigenvalues and eigenvectors of a general symmetric block circulant matrix are characterized. We apply this result in our situation, when $n = 2k + 1$ is odd. In view of this, the spectrum of $\mathfrak{C}_n^{(s)}$ is the union of spectra of the matrices

$$\mathbf{M}_j = \mathbf{C}(0) + \sum_{h=1}^k [\mathbf{C}(h) \rho_j^h + \mathbf{C}^T(h) \rho_j^{-h}] = \mathbf{C}(0) + \sum_{h=1}^k [\mathbf{C}(h) e^{i\omega_j h} + \mathbf{C}^T(h) e^{-i\omega_j h}]$$

for $j = 0, 1, \dots, n-1$; whereas, the eigenvectors are obtained by compounding the eigenvectors of these $d \times d$ matrices. So we need the spectral decomposition of the matrices $\mathbf{M}_0 = \mathbf{C}(0) + \sum_{h=1}^k [\mathbf{C}(h) + \mathbf{C}^T(h)]$ and

$$\mathbf{M}_j = \mathbf{C}(0) + \sum_{h=1}^k [(\mathbf{C}(h) + \mathbf{C}^T(h)) \cos(\omega_j h) + i(\mathbf{C}(h) - \mathbf{C}^T(h)) \sin(\omega_j h)]$$

for $j = 1, 2, \dots, n-1$. Since $\mathbf{C}(h) + \mathbf{C}^T(h)$ is symmetric and $\mathbf{C}(h) - \mathbf{C}^T(h)$ is anti-symmetric with 0 diagonal, \mathbf{M}_j is self-adjoint for each j and has real eigenvalues with corresponding orthonormal set of eigenvectors of possibly complex coordinates. Indeed, \mathbf{M}_j may have complex entries if $j \neq 0$; actually, $\sum_{h=1}^k (\mathbf{C}(h) + \mathbf{C}^T(h)) \cos(\omega_j h)$ is the real and $\sum_{h=1}^k (\mathbf{C}(h) - \mathbf{C}^T(h)) \sin(\omega_j h)$ is the imaginary part of \mathbf{M}_j .

It is easy to see that $\mathbf{M}_{n-j} = \overline{\mathbf{M}_j}$ (entrywise conjugate), therefore, it has the same (real) eigenvalues as \mathbf{M}_j , but its (complex) eigenvectors are the (componentwise) complex conjugates of the eigenvectors of \mathbf{M}_j . We also need the following form of this matrix:

$$\begin{aligned} \mathbf{M}_{n-j} &= \mathbf{C}(0) + \sum_{h=1}^k [(\mathbf{C}(h) + \mathbf{C}^T(h)) \cos(\omega_j h) - i(\mathbf{C}(h) - \mathbf{C}^T(h)) \sin(\omega_j h)] \\ &= \mathbf{C}(0) + \sum_{h=1}^k [\mathbf{C}(h)e^{-i\omega_j h} + \mathbf{C}^T(h)e^{i\omega_j h}], \quad j = 1, \dots, n-1. \end{aligned} \quad (3)$$

Summarizing, for odd $n = 2k+1$, the nd eigenvalues of $\mathfrak{C}_n^{(s)}$ are obtained as the union of the (real) eigenvalues of \mathbf{M}_0 and those of \mathbf{M}_j ($j = 1, \dots, k$) duplicated. Note that for even n , similar arguments hold with the difference that there the spectrum of $\mathfrak{C}_n^{(s)}$ is the union of the eigenvalues of \mathbf{M}_0 and \mathbf{M}_{n-1} , whereas the eigenvalues of $\mathbf{M}_1, \dots, \mathbf{M}_{\frac{n}{2}-1}$ are duplicated.

The eigenvectors of $\mathfrak{C}_n^{(s)}$ are obtainable by compounding the d (usually complex) orthonormal eigenvectors of the $d \times d$ self-adjoint matrices $\mathbf{M}_0, \mathbf{M}_1, \dots, \mathbf{M}_{n-1}$ as follows. For $j = 1, \dots, k$: if \mathbf{v} is a unit-length eigenvector of \mathbf{M}_j with eigenvalue λ , then in [3] it is proved that the compound vector

$$(\mathbf{v}^T, \rho_j \mathbf{v}^T, \rho_j^2 \mathbf{v}^T, \dots, \rho_j^{n-1} \mathbf{v}^T)^T \in \mathbb{C}^{nd}$$

is an eigenvector of $\mathfrak{C}_n^{(s)}$ with the same eigenvalue λ . It has squared norm

$$\mathbf{v}^* \mathbf{v} (1 + \rho_j \rho_j^{-1} + \rho_j^2 \rho_j^{-2} + \dots + \rho_j^{n-1} \rho_j^{-(n-1)}) = n.$$

Therefore, the vector

$$\mathbf{w} = \frac{1}{\sqrt{n}} (\mathbf{v}^T, \rho_j \mathbf{v}^T, \rho_j^2 \mathbf{v}^T, \dots, \rho_j^{n-1} \mathbf{v}^T)^T \in \mathbb{C}^{nd} \quad (4)$$

is a unit-norm eigenvector (of complex coordinates) of $\mathfrak{C}_n^{(s)}$.

Further, if

$$\mathbf{z} = \frac{1}{\sqrt{n}} (\mathbf{t}^T, \rho_\ell \mathbf{t}^T, \rho_\ell^2 \mathbf{t}^T, \dots, \rho_\ell^{n-1} \mathbf{t}^T)^T \in \mathbb{C}^{nd}$$

is another unit-norm eigenvector of $\mathfrak{C}_n^{(s)}$ compounded from a unit-norm eigenvector \mathbf{t} of another \mathbf{M}_ℓ ($\ell \neq j$), then \mathbf{w} and \mathbf{z} are orthogonal, irrespective whether \mathbf{M}_ℓ has the same eigenvalue λ as \mathbf{M}_j or not. Similar construction holds starting with the eigenvectors of \mathbf{M}_0 .

Here for each $j = 0, 1, \dots, n-1$, there are d pairwise orthonormal eigenvectors (potential \mathbf{v} s) of \mathbf{M}_j , and the so obtained \mathbf{w} s are also pairwise orthonormal. Assume that the eigenvectors of \mathbf{M}_j are enumerated in non-increasing order of its (real) eigenvalues, and the inflated \mathbf{w} s also follow this ordering, for $j = 0, 1, \dots, n-1$.

Choose a unit-norm eigenvector $\mathbf{v} \in \mathbb{C}^d$ of \mathbf{M}_j with (real) eigenvalue λ . Then $\bar{\mathbf{v}} \in \mathbb{C}^d$ is the corresponding unit-norm eigenvector of \mathbf{M}_{n-j} with the same eigenvalue λ . Consider the compounded $\mathbf{w} \in \mathbb{C}^{nd}$ and $\bar{\mathbf{w}} \in \mathbb{C}^{nd}$ obtained from them by (4). We learned that they are orthonormal eigenvectors of $\mathfrak{C}_n^{(s)}$ corresponding to the eigenvalue λ with multiplicity (at least) two. From them, corresponding to this double eigenvalue λ , the new orthonormal pair of eigenvectors

$$\frac{\mathbf{w} + \bar{\mathbf{w}}}{\sqrt{2}} \quad \text{and} \quad -i \frac{\mathbf{w} - \bar{\mathbf{w}}}{\sqrt{2}} \quad (5)$$

is constructed, but they, in this order, occupy the original positions of \mathbf{w} and $\bar{\mathbf{w}}$. They have real coordinates and unit norm. (Actually, their coordinates are the $\sqrt{2}$ multiples of the real and imaginary parts of the coordinates of \mathbf{w} .) It is in accord with the fact that a real symmetric matrix, as $\mathfrak{C}_n^{(s)}$, must have an orthogonal system of eigenvectors with real coordinates too. We do not go in details, neither discuss defective cases.

Consider $\mathbf{u}_1, \dots, \mathbf{u}_{nd}$, the so obtained orthonormal set of eigenvectors (of real coordinates) of $\mathbf{C}_n^{(s)}$ (in the above ordering), and denote by $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_{nd})$ the $nd \times nd$ (real) orthogonal matrix containing them columnwise. Let

$$\mathbf{C}_n^{(s)} = \mathbf{U}\mathbf{\Lambda}^{(s)}\mathbf{U}^T \quad (6)$$

be the corresponding spectral decomposition. After this preparation, we are able to prove the following theorem.

3 The main theorem

Theorem 1. *Let $\{\mathbf{X}_t\}$ be d -dimensional weakly stationary time series of real components. Denoting by $\mathbf{C}(h) = [c_{ij}(h)]$ the $d \times d$ autocovariance matrices ($\mathbf{C}(-h) = \mathbf{C}^T(h)$, $h \in \mathbb{Z}$) in the time domain, assume that their entries are absolutely summable, i.e., $\sum_{h=0}^{\infty} |c_{pq}(h)| < \infty$ for $p, q = 1, \dots, d$. Then, the self-adjoint, positive semidefinite spectral density matrix \mathbf{f} exists in the frequency domain, and it is defined by*

$$\mathbf{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbf{C}(h)e^{-ih\omega}, \quad \omega \in [0, 2\pi].$$

For odd $n = 2k + 1$, consider $\mathbf{X}_1, \dots, \mathbf{X}_n$ and the block Toeplitz matrix \mathbf{C}_n of (1); further, the Fourier frequencies $\omega_j = \frac{2\pi j}{n}$ for $j = 0, \dots, n - 1$. Let \mathbf{D}_n be the $dn \times dn$ diagonal matrix that contains the spectra of the matrices $\mathbf{f}(0), \mathbf{f}(\omega_1), \mathbf{f}(\omega_2), \dots, \mathbf{f}(\omega_k), \mathbf{f}(\omega_k), \dots, \mathbf{f}(\omega_2), \mathbf{f}(\omega_1)$ in its main diagonal, i.e.,

$$\mathbf{D}_n = \text{diag}(\text{spec } \mathbf{f}(0), \text{spec } \mathbf{f}(\omega_1), \dots, \text{spec } \mathbf{f}(\omega_k), \text{spec } \mathbf{f}(\omega_k), \dots, \text{spec } \mathbf{f}(\omega_1)).$$

Here *spec* contains the eigenvalues of the affected matrix in non-increasing order if not otherwise stated. (The duplication is due to the fact that $\mathbf{f}(\omega_j) = \mathbf{f}(\omega_{n-j})$, $j = 1, \dots, k$, for real time series.) Then, with the spectral decomposition (6),

$$\mathbf{U}^T \mathbf{C}_n \mathbf{U} - 2\pi \mathbf{D}_n \rightarrow \mathbf{O}, \quad n \rightarrow \infty,$$

i.e., the entries of the matrix $\mathbf{U}^T \mathbf{C}_n \mathbf{U} - 2\pi \mathbf{D}_n$ tend to 0 uniformly as $n \rightarrow \infty$.

Proof. We saw that $\mathbf{U}^T \mathbf{C}_n^{(s)} \mathbf{U} = \mathbf{\Lambda}^{(s)}$. Recall that the eigenvalues in the diagonal of $\mathbf{\Lambda}^{(s)}$ comprise the union of spectra of the matrices \mathbf{M}_0 and those of $\mathbf{M}_1, \dots, \mathbf{M}_{n-1}$, which are the same as the eigenvalues of \mathbf{M}_0 and those of $\mathbf{M}_{n-1}, \dots, \mathbf{M}_{n-k}$ of (3), duplicated. But these matrices are finite sub-sums (for $|h| \leq k$) of the infinite summations

$$2\pi \mathbf{f}(\omega_j) = \sum_{h=-\infty}^{\infty} \mathbf{C}(h)e^{-ih\omega_j} = \mathbf{C}(0) + \sum_{h=1}^{\infty} [\mathbf{C}(h)e^{-i\omega_j h} + \mathbf{C}^T(h)e^{i\omega_j h}].$$

So, by the absolute summability of the autocovariances, and because the eigenvalues depend continuously on the underlying matrices, the pairwise distances between the eigenvalues of \mathbf{M}_j and the corresponding eigenvalues of $2\pi \mathbf{f}(\omega_j)$ (both in non-increasing order) tend to 0 as $n \rightarrow \infty$, for $j = 0, 1, \dots, k$. Indeed, the absolute summability of the entries of $\mathbf{C}(h)$ s implies that the diagonal entries of the diagonal matrix $\mathbf{\Lambda}^{(s)} - 2\pi \mathbf{D}_n$ are bounded in absolute value by

$$\max_{p,q \in \{1, \dots, d\}} \sum_{|h| > k} |c_{pq}(h)| \rightarrow 0, \quad n = 2k + 1 \rightarrow \infty.$$

Consequently, the matrix $\mathbf{\Lambda}^{(s)} - 2\pi \mathbf{D}_n$ tends to the zero matrix entrywise uniformly as $n \rightarrow \infty$. Therefore, it remains to show that the entries of $\mathbf{U}^T \mathbf{C}_n \mathbf{U} - \mathbf{U}^T \mathbf{C}_n^{(s)} \mathbf{U}$ tend to 0 uniformly as $n \rightarrow \infty$.

Before doing this, some facts should be clarified.

- The p th row sum of \mathbf{M}_j is bounded by

$$\sum_{q=1}^d |c_{pq}(0)| + \sum_{q=1}^d \sum_{h=1}^k |c_{pq}(h)| + \sum_{q=1}^d \sum_{h=1}^k |c_{qp}(h)| \leq d c_{pp}(0) + 2dL,$$

for $p \in \{1, \dots, d\}$ with $L = \max_{p,q \in \{1, \dots, d\}} \sum_{h=1}^{\infty} |c_{pq}(h)| > 0$, independently of n , because of the absolute summability of the entries of $\mathbf{C}(h)$. This is true for any $j \in \{0, 1, \dots, n-1\}$. For simplicity, consider (any) one of the \mathbf{M}_j s, and denote it by $\mathbf{M} = [m_{pq}]_{p,q=1}^d$. Then

$$\|\mathbf{M}\|_{\infty} = \max_{p \in \{1, \dots, d\}} \sum_{q=1}^d |m_{pq}| \leq d \max_{p \in \{1, \dots, d\}} c_{pp}(0) + 2dL = K.$$

As the spectral radius of \mathbf{M} is at most $\|\mathbf{M}\|_{\infty}$, any eigenvalue λ of \mathbf{M} is bounded in absolute value by K (independently of n).

- Recall that \mathbf{u} is compounded via (4) and (5) from the primitive roots. Therefore, its coordinates are bounded by $\sqrt{\frac{2}{n}}$ in absolute value.

Now we are ready to show that

$$|\mathbf{u}_i^T \mathfrak{C}_n^{(s)} \mathbf{u}_j - \mathbf{u}_i^T \mathfrak{C}_n \mathbf{u}_j| \rightarrow 0, \quad n \rightarrow \infty$$

uniformly in $i, j \in \{1, \dots, nd\}$. Recall that in the $nd \times nd$ matrices $\mathfrak{C}_n^{(s)}$ and \mathfrak{C}_n the (m, ℓ) blocks are the same if $|m-\ell| \leq k$. Denote by $\mathbf{u}_{i,m}$ and $\mathbf{u}_{j,\ell}$ the m th and ℓ th blocks of the unit-norm eigenvectors \mathbf{u}_i and \mathbf{u}_j , respectively. Then

$$\begin{aligned} |\mathbf{u}_i^T (\mathfrak{C}_n^{(s)} - \mathfrak{C}_n) \mathbf{u}_j| &= \left| \sum_{m=1}^k \sum_{\ell=1}^m [\mathbf{u}_{i,\ell}^T (\mathbf{C}(m) - \mathbf{C}(n-m)) \mathbf{u}_{j,n-m+\ell} + \mathbf{u}_{i,n-m+\ell}^T (\mathbf{C}(n-m) - \mathbf{C}(m)) \mathbf{u}_{j,\ell}] \right| \\ &\leq \left| \sqrt{\frac{2}{n}} \sum_{m=1}^k \mathbf{1}_d^T (\mathbf{C}(m) - \mathbf{C}(n-m)) \sum_{\ell=1}^m \mathbf{u}_{j,n-m+\ell} \right| \\ &\quad + \left| \sqrt{\frac{2}{n}} \sum_{m=1}^k \sum_{\ell=1}^m \mathbf{u}_{i,n-m+\ell}^T (\mathbf{C}(m) - \mathbf{C}(n-m)) \mathbf{1}_d \right| \\ &\leq 2 \sqrt{\frac{2}{n}} \sqrt{\frac{2}{n}} \left| \sum_{m=1}^k m \mathbf{1}_d^T (\mathbf{C}(m) - \mathbf{C}(n-m)) \mathbf{1}_d \right| \\ &\leq \frac{4}{n} \left(\sum_{m=1}^k m \sum_{p=1}^d \sum_{q=1}^d |c_{pq}(m)| + \sum_{m=1}^k m \sum_{p=1}^d \sum_{q=1}^d |c_{pq}(n-m)| \right) \\ &\leq 4d^2 \left(\max_{p,q \in \{1, \dots, d\}} \sum_{m=1}^k \frac{m}{n} |c_{pq}(m)| + \max_{p,q \in \{1, \dots, d\}} \sum_{m=1}^k \frac{m}{n} |c_{pq}(n-m)| \right) \\ &\leq 4d^2 \left(\max_{p,q \in \{1, \dots, d\}} \sum_{m=1}^k \frac{m}{n} |c_{pq}(m)| + \max_{p,q \in \{1, \dots, d\}} \sum_{m=n-k}^{n-1} \frac{k}{n} |c_{pq}(m)| \right), \end{aligned}$$

where $\mathbf{1}_d \in \mathbb{R}^d$ is the vector of all 1 coordinates and so, the quadratic form $\mathbf{1}_d^T (\mathbf{C}(m) - \mathbf{C}(n-m)) \mathbf{1}_d$ is the sum of the entries of $\mathbf{C}(m) - \mathbf{C}(n-m)$. In the last line, the second term converges to 0, since it is bounded by $\sum_{m=k}^{\infty} |c_{pq}(m)|$ (indeed, $\sum_{m=n-k}^{n-1} \frac{k}{n} |c_{pq}(m)| \leq \sum_{m=k}^{\infty} |c_{pq}(m)|$ as $k < n-k$), and together with n , k tends to ∞ too; further, it holds uniformly for all $p, q \in \{1, \dots, d\}$. The first term for every p, q pair also tends to 0 as $n \rightarrow \infty$ by the discrete version of the dominated convergence theorem (for series), see the forthcoming Lemma 1. Indeed, the summand is dominated by $|c_{pq}(m)|$ and $\sum_{m=1}^{\infty} |c_{pq}(m)| < \infty$; further, $\frac{m}{n} |c_{pq}(m)| \rightarrow 0$ as $n \rightarrow \infty$, for any fixed m . Consequently, $\sum_{m=1}^{\infty} \frac{m}{n} |c_{pq}(m)|$ tends to 0, and so does $\sum_{m=1}^k \frac{m}{n} |c_{pq}(m)|$ as $n \rightarrow \infty$. It holds uniformly for all p, q , and also for all i, j , so the proof is complete. \square

Lemma 1 (Dominated convergence theorem for sums, discrete version). *Consider $\sum_{m=1}^{\infty} f_n(m)$ and assume that $|f_n(m)| \leq g(m)$ with $\sum_{m=1}^{\infty} g(m) < \infty$. If $\lim_{n \rightarrow \infty} f_n(m) = f(m)$ exists $\forall m \in \mathbb{N}$, then*

$$\lim_{n \rightarrow \infty} \sum_{m=1}^{\infty} f_n(m) = \sum_{m=1}^{\infty} f(m).$$

The proof of Theorem 1 suggests that the speed of the convergence depends on how rapidly the tail of the absolutely summable autocovariances tends to zero. For example, in case of VARMA (vector autoregressive) processes the autocovariances decrease exponentially, and this implies the convergence at a geometric rate.

4 Relations and consequences of the main theorem

4.1 Bounds for the eigenvalues of \mathfrak{C}_n

Analogously to the 1D statement (see [1]), for the eigenvalues of the block Toeplitz matrix \mathfrak{C}_n the following lower and upper bounds are obtainable.

Proposition 1. *Assume that for the spectra of the spectral densities \mathbf{f} of the d -dimensional weakly stationary process $\{\mathbf{X}_t\}$ of real coordinates the following hold:*

$$m := \inf_{\omega \in [0, 2\pi], q \in \{1, \dots, d\}} \lambda_q(\mathbf{f}(\omega)) > 0,$$

$$M := \sup_{\omega \in [0, 2\pi], q \in \{1, \dots, d\}} \lambda_q(\mathbf{f}(\omega)) < \infty.$$

(Note that under the conditions of Theorem 1, $\mathbf{f}(\omega) > 0$ and it is continuous almost everywhere on $[0, 2\pi]$, so the above conditions are readily satisfied.)

Then for the eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{nd}$ of the block Toeplitz matrix \mathfrak{C}_n the following holds:

$$2\pi m \leq \lambda_1 \leq \lambda_{nd} \leq 2\pi M.$$

Proof. Let λ be an arbitrary eigenvalue of \mathfrak{C}_n with a corresponding eigenvector $\mathbf{x} \in \mathbb{C}^{nd}$ (i.e., $\mathfrak{C}_n \mathbf{x} = \lambda \mathbf{x}$), and denote by \mathbf{x}_j the j th block of it ($j = 1, \dots, n$). Take the spectral decomposition of the spectral density matrix \mathbf{f} :

$$\mathbf{f}(\omega) = \sum_{\ell=1}^d \lambda_\ell(\mathbf{f}(\omega)) \cdot \mathbf{u}_\ell(\mathbf{f}(\omega)) \cdot \mathbf{u}_\ell^*(\mathbf{f}(\omega)).$$

Then we can write that

$$\begin{aligned} \lambda |\mathbf{x}|^2 &= \lambda \mathbf{x}^* \mathbf{x} = \mathbf{x}^* \mathfrak{C}_n \mathbf{x} \\ &= \mathbf{x}^* \cdot \int_{-\pi}^{\pi} \left[e^{-i(j-k)\omega} \mathbf{f}(\omega) \right]_{j,k=1}^n d\omega \cdot \mathbf{x} \\ &= \int_{-\pi}^{\pi} \sum_{j,k=1}^n e^{-i(j-k)\omega} \mathbf{x}_j^* \mathbf{f}(\omega) \mathbf{x}_k d\omega \\ &= \int_{-\pi}^{\pi} \sum_{j,k=1}^n e^{-i(j-k)\omega} \sum_{\ell=1}^d \lambda_\ell(\mathbf{f}(\omega)) \cdot \mathbf{x}_j^* \cdot \mathbf{u}_\ell(\mathbf{f}(\omega)) \cdot \mathbf{u}_\ell^*(\mathbf{f}(\omega)) \cdot \mathbf{x}_k d\omega \\ &= \int_{-\pi}^{\pi} \sum_{\ell=1}^d \lambda_\ell(\mathbf{f}(\omega)) \sum_{j,k=1}^n e^{-ij\omega} \mathbf{x}_j^* \cdot \mathbf{u}_\ell(\mathbf{f}(\omega)) \cdot \mathbf{u}_\ell^*(\mathbf{f}(\omega)) \cdot \mathbf{x}_k \cdot e^{ik\omega} d\omega \\ &= \int_{-\pi}^{\pi} \sum_{\ell=1}^d \lambda_\ell(\mathbf{f}(\omega)) \left| \sum_{j=1}^n e^{-ij\omega} \cdot \mathbf{x}_j^* \cdot \mathbf{u}_\ell(\mathbf{f}(\omega)) \right|^2 d\omega \\ &\leq M \sum_{j,k=1}^n \mathbf{x}_j^* \cdot \int_{-\pi}^{\pi} e^{-i(j-k)\omega} \sum_{\ell=1}^d \mathbf{u}_\ell(\mathbf{f}(\omega)) \cdot \mathbf{u}_\ell^*(\mathbf{f}(\omega)) d\omega \cdot \mathbf{x}_k \end{aligned}$$

$$= 2\pi M \sum_{j=1}^n \mathbf{x}_j^* \mathbf{x}_j = 2\pi M |\mathbf{x}|^2.$$

This proves that $\lambda \leq 2\pi M$ for any eigenvalue of \mathfrak{C}_n . The proof of the fact that $\lambda \geq 2\pi m$ is similar. \square

4.2 PC transformation as discrete Fourier transformation

The complex PC transform of the collection of random vectors $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)^T$ of real coordinates is the random vector $\mathbf{Z} = (\mathbf{Z}_1^T, \dots, \mathbf{Z}_n^T)^T$ of complex coordinates obtained by

$$\mathbf{Z} = \mathbf{W}^* \mathbf{X}.$$

Here, analogously to (6), $\mathfrak{C}_n^{(s)}$ has the spectral decomposition

$$\mathfrak{C}_n^{(s)} = \mathbf{W} \boldsymbol{\Lambda}^{(s)} \mathbf{W}^* \quad (7)$$

with complex eigenvectors too: $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_{nd})$, where \mathbf{w}_s are the orthonormal eigenvectors of $\mathfrak{C}_n^{(s)}$ (of complex coordinates), introduced in Section 2.

To relate the PC transformation to a discrete Fourier transformation, we also make PC transformations within the blocks. For this purpose we use the eigenvectors in the columns of \mathbf{W} (of complex coordinates) in the ordering described in Section 2. We utilize their block structure and also assume that they are already normalized to have a complete orthonormal system in \mathbb{C}^{nd} .

By Theorem 1, $\mathbb{E} \mathbf{Z} \mathbf{Z}^* \sim 2\pi \mathbf{D}_n$, so the coordinates of \mathbf{Z} are asymptotically uncorrelated, for ‘large’ n . Instead, we consider the blocks \mathbf{Z}_j s of it, and perform a ‘partial principal component transformation’ (in d -dimension) of them. Let $\mathbf{w}_{1j}, \dots, \mathbf{w}_{dj}$ be the columns of \mathbf{W} corresponding to the coordinates of \mathbf{Z}_j . In view of (4), \mathbf{Z}_j can be written as

$$\mathbf{Z}_j = \frac{1}{\sqrt{n}} (\mathbf{V}_j^* \otimes \mathbf{r}^*) \mathbf{X},$$

where $\mathbf{r}^* = (1, \rho_j^{-1}, \rho_j^{-2}, \dots, \rho_j^{-(n-1)})$ and \mathbf{V}_j is the $d \times d$ unitary matrix in the spectral decomposition $\mathbf{M}_j = \mathbf{V}_j \boldsymbol{\Lambda}_j \mathbf{V}_j^*$. Because of $\mathbb{E} \mathbf{Z}_j \mathbf{Z}_j^* = \boldsymbol{\Lambda}_j$ (apparently from the proof of Theorem 1), we have that

$$\mathbb{E} (\mathbf{V}_j \mathbf{Z}_j) (\mathbf{V}_j \mathbf{Z}_j)^* = \mathbf{V}_j \boldsymbol{\Lambda}_j \mathbf{V}_j^* = \mathbf{M}_j.$$

At the same time,

$$\mathbf{V}_j \mathbf{Z}_j = \frac{1}{\sqrt{n}} \mathbf{V}_j (\mathbf{V}_j^* \otimes \mathbf{r}^*) \mathbf{X} = \frac{1}{\sqrt{n}} (\mathbf{I}_d \otimes \mathbf{r}^*) \mathbf{X} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{X}_t e^{-it\omega_j}, \quad j = 1, \dots, n.$$

This is the discrete Fourier transform of $\mathbf{X}_1, \dots, \mathbf{X}_n$. It is in accord with the existence of the orthogonal increment process $\{\mathbf{Z}_\omega\}$ (see [1]) of which $\mathbf{V}_j \mathbf{Z}_j \sim \mathbf{Z}_{\omega_j} - \mathbf{Z}_{\omega_{j-1}}$ is the discrete analogue. Note that a weakly stationary time series itself can be represented (with probability one) as

$$\mathbf{X}_t = \int_{-\pi}^{\pi} e^{it\omega} d\mathbf{Z}_\omega, \quad t \in \mathbb{Z}$$

with the process \mathbf{Z}_ω of orthogonal increments, where $\mathbf{F}(\omega) = \mathbb{E}(\mathbf{Z}_\omega \mathbf{Z}_\omega^*)$, $\mathbf{F}(\omega) = \int_{-\pi}^{\omega} \mathbf{f}(\theta) d\theta$, $\omega \in [-\pi, \pi]$. This is sometimes called Cramér representation. (It generalizes the case of the superposition of sinusoids, where the process has point spectrum.)

Also, $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ are asymptotically pairwise orthogonal akin to $\mathbf{V}_1 \mathbf{Z}_1, \dots, \mathbf{V}_n \mathbf{Z}_n$. Further,

$$\mathbb{E} (\mathbf{V}_j \mathbf{Z}_j) (\mathbf{V}_j \mathbf{Z}_j)^* \sim 2\pi \mathbf{f}(\omega_j),$$

and it is in accord with the fact that

$$\mathbb{E} \mathbf{Z}_j \mathbf{Z}_j^* \sim 2\pi \text{diag spec } \mathbf{f}(\omega_j),$$

for $j = 0, 1, \dots, n-1$ when n is ‘large’.

4.3 Best low-rank approximation in the frequency domain, and low-dimensional approximation in the time domain

Let $\{\mathbf{X}_t\}_{t=1}^n$ be the finite part of a d -dimensional process of real coordinates and constant rank $1 \leq r \leq d$. Its discrete Fourier transform, discussed in Section 4.2, is

$$\mathbf{T}_j = \mathbf{V}_j \mathbf{Z}_j = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{X}_t e^{-it\omega_j}, \quad j = 0, \dots, n-1.$$

More precisely, $\mathbf{T}_0 = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{X}_t$,

$$\mathbf{T}_j = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{X}_t [\cos(t\omega_j) - i \sin(t\omega_j)],$$

and $\mathbf{T}_{n-j} = \overline{\mathbf{T}}_j$, for $j = 1, \dots, k$ ($n = 2k + 1$). Therefore,

$$\mathbf{Z}_j = \mathbf{V}_j^{-1} \mathbf{T}_j = \mathbf{V}_j^* \mathbf{T}_j, \quad j = 0, \dots, n-1.$$

It can easily be seen that $\mathbf{Z}_{n-j} = \overline{\mathbf{Z}}_j$.

To find the best m -rank approximation ($1 < m \leq r$) of the process, we project the d -dimensional vector \mathbf{T}_j onto the subspace spanned by the m leading eigenvectors of \mathbf{V}_j . Important that the eigenvalues in Λ_j are in non-increasing order. Let us denote the eigenvectors corresponding to the m largest eigenvalues by $\mathbf{v}_{j1}, \dots, \mathbf{v}_{jm}$. Then

$$\widehat{\mathbf{T}}_j := \text{Proj}_{\text{Span}\{\mathbf{v}_{j1}, \dots, \mathbf{v}_{jm}\}} \mathbf{T}_j = \sum_{\ell=1}^m (\mathbf{v}_{j\ell}^* \mathbf{V}_j \mathbf{Z}_j) \mathbf{v}_{j\ell} = \sum_{\ell=1}^m Z_{j\ell} \mathbf{v}_{j\ell},$$

and $\widehat{\mathbf{T}}_{n-j} = \overline{\widehat{\mathbf{T}}}_j$, for $j = 1, \dots, k$ (by the previous considerations), where $n = 2k + 1$. Further,

$$\widehat{\mathbf{T}}_0 := \sum_{\ell=1}^m Z_{0\ell} \mathbf{v}_{0\ell}.$$

So, for each j , the resulting vector is the linear combination of the vectors $\mathbf{v}_{j\ell}$ s with the corresponding coordinates $Z_{j\ell}$ s of \mathbf{Z}_j , $\ell = 1, \dots, m$.

Eventually, we find the m -rank approximation of \mathbf{X}_t by inverse Fourier transformation:

$$\begin{aligned} \widehat{\mathbf{X}}_t &:= \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \widehat{\mathbf{T}}_j e^{it\omega_j} = \\ &= \frac{1}{\sqrt{n}} \left\{ \widehat{\mathbf{T}}_0 + \sum_{j=1}^k [(\widehat{\mathbf{T}}_j + \overline{\widehat{\mathbf{T}}}_j) \cos(t\omega_j) + i(\widehat{\mathbf{T}}_j - \overline{\widehat{\mathbf{T}}}_j) \sin(t\omega_j)] \right\} \\ &= \frac{1}{\sqrt{n}} \left\{ \widehat{\mathbf{T}}_0 + \sum_{j=1}^k [(2\text{Re}(\widehat{\mathbf{T}}_j) \cos(t\omega_j) + i \cdot 2i \cdot \text{Im}(\widehat{\mathbf{T}}_j) \sin(t\omega_j)] \right\} \\ &= \frac{1}{\sqrt{n}} \left\{ \widehat{\mathbf{T}}_0 + 2 \sum_{j=1}^k [\text{Re}(\widehat{\mathbf{T}}_j) \cos(t\omega_j) - \text{Im}(\widehat{\mathbf{T}}_j) \sin(t\omega_j)] \right\}. \end{aligned}$$

Apparently, the vectors $\widehat{\mathbf{X}}_t$ ($t = 1, \dots, n$) all have real coordinates ($n = 2k + 1$).

In this way, we have a lower rank process with spectral density of rank $m \leq r$. Note that if the process is regular (e.g., it has a rational spectral density), then so is its low-rank approximation.

Note that in the static PCA (e.g., [5]), the uncorrelated PCs explain the largest possible proportion of the total variance of the (usually correlated) components of a random vector, their variances are in decreasing order and those are the eigenvalues of the covariance matrix. Even more is true: for any m , which is at most

the dimension of the random vector, the first m principal components provide the best m -rank approximation of the random vector in mean square. This theory also guarantees that the ‘larger’ the gap between the m th and $(m + 1)$ th eigenvalues (in non-increasing order) of the spectral density matrix, the ‘smaller’ the approximation error is. Unlike the static PCA, the complex PCs, i.e. \mathbf{Z}_j s in the frequency domain make rise to explain proportions of the complete data variance both cross-sectionally and longitudinally (briefly, vertically and horizontally) as follows.

Consider the spectral decomposition of the $d \times d$ matrix $\mathbf{M}_j = \mathbf{V}_j \mathbf{\Lambda}_j \mathbf{V}_j^*$, and denote by $\lambda_j^{(1)} \geq \dots \geq \lambda_j^{(d)}$ the eigenvalues of \mathbf{M}_j , $j = 1, \dots, k$ ($n = 2k + 1$). At frequency ω_j . The eigenvalue $\lambda_j^{(\ell)}$ is the contribution of the ℓ th PC to the variance of the increment process at frequency ω_j . Let

$$\lambda_{cum}^{(\ell)} := \sum_{j=1}^k \lambda_j^{(\ell)}, \quad \ell = 1, \dots, m$$

be the cumulative contribution of the ℓ th PC to the total variance. The relative contribution of the first m PCs to the total variance is

$$\frac{\sum_{\ell=1}^m \lambda_{cum}^{(\ell)}}{\sum_{\ell=1}^d \lambda_{cum}^{(\ell)}}. \quad (8)$$

Then the number m^* of the PCs retained is the smallest index for which the above relative contribution exceeds a certain percent, or else, when it increases drastically. This is the *vertical* contribution.

As for the *horizontal* contribution, we investigate the contribution of the Fourier frequencies to the total variance. For $\ell = 1, \dots, m$: let $\omega_{j^*}^{(\ell)}$ denote the argmax of the first local maximum of $\lambda_j^{(\ell)}$. Then the corresponding eigenvector of \mathbf{M}_{j^*} , i.e. the ℓ th column of \mathbf{V}_{j^*} contains the (complex) factor loadings of the ℓ th PC at that prominent frequency. Based on the absolute values (which is calculated from the the real and imaginary parts of these coordinates) we can explain the ℓ th PC, by telling which variables (with the largest loadings) contribute to it most significantly.

To back-transform the PC process into the time domain, note that

$$\mathbf{Z}_{j\ell} = \mathbf{v}_{j\ell}^* \mathbf{T}_j, \quad \ell = 1, \dots, m$$

defines the coordinates of an m -dimensional approximation of \mathbf{T}_j , $m \leq r \leq d$. This is the m -dimensional vector $\tilde{\mathbf{T}}_j = (Z_{j1}, \dots, Z_{jm})^T$. I.e., we take the first m complex PCs in each block (it is important that the entries in the diagonal of each $\mathbf{\Lambda}_j$ are in non-increasing order).

Then the m -dimensional approximation of \mathbf{X}_t by the PC process is as follows:

$$\begin{aligned} \tilde{\mathbf{X}}_t &:= \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \tilde{\mathbf{T}}_j e^{it\omega_j} = \\ &= \frac{1}{\sqrt{n}} \left\{ \tilde{\mathbf{T}}_0 + \sum_{j=1}^k [(\tilde{\mathbf{T}}_j + \bar{\tilde{\mathbf{T}}}_j) \cos(t\omega_j) + i(\tilde{\mathbf{T}}_j - \bar{\tilde{\mathbf{T}}}_j) \sin(t\omega_j)] \right\} \\ &= \frac{1}{\sqrt{n}} \left\{ \tilde{\mathbf{T}}_0 + \sum_{j=1}^k [(2\operatorname{Re}(\tilde{\mathbf{T}}_j) \cos(t\omega_j) + i \cdot 2i \cdot \operatorname{Im}(\tilde{\mathbf{T}}_j) \sin(t\omega_j))] \right\} \\ &= \frac{1}{\sqrt{n}} \left\{ \tilde{\mathbf{T}}_0 + 2 \sum_{j=1}^k [\operatorname{Re}(\tilde{\mathbf{T}}_j) \cos(t\omega_j) - \operatorname{Im}(\tilde{\mathbf{T}}_j) \sin(t\omega_j)] \right\} \end{aligned}$$

that again results in real coordinates. Equivalently, the m -dimensional PC process is:

$$\tilde{\mathbf{X}}_t = \frac{1}{\sqrt{n}} \left(\sum_{j=0}^{n-1} Z_{j1} e^{it\omega_j}, \dots, \sum_{j=0}^{n-1} Z_{jm} e^{it\omega_j} \right)^T. \quad (9)$$

4.4 Numerical example

The previously detailed low-rank approximation is illustrated on a financial dataset [6] containing stock exchange returns: Istanbul stock exchange national 100 index (ISE), S&P 500 return index (SP), stock market return index of Germany (DAX), UK (FTSE), Japan (NIKKEI), Brazil (BOVESPA), the MSCI European index (EU), and the MSCI emerging markets index (EM); ranging from Jan 5, 2009 to Feb 22, 2011. It is a $d = 8$ dimensional time-series dataset that we studied up to length $n = 535$. The days for which the Turkish stock exchange was closed are excluded, and in case of missing data for the other indices, the previous day's value was imputed, see [6].

In Figure 1, the three largest eigenvalue processes of the estimated M_j matrices are shown in the frequency domain, the others were close to zero. Based on this, the time-series is (approximately) of rank 3, thus we can apply the outlined low-rank approximation with $m = 3$. In Figure 2, the rank 3 approximation of the individual variables and the original data is illustrated. There are the calculated RMSE values under each subplot. The 3 leading PCs, back-transformed to the time domain, are to be found in Figure 3.

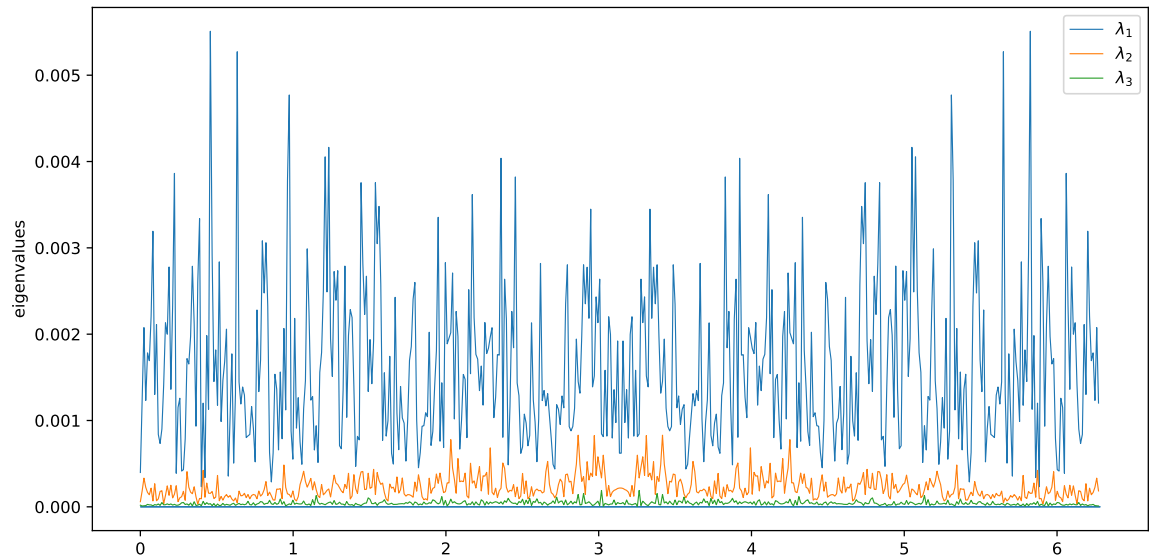


Figure 1: The three largest eigenvalue processes of the estimated M_j ($j = 0, \dots, 534$) matrices over $[0, 2\pi]$ (symmetric with respect to π), ordered decreasingly.

However, according to Equation (8), the first two PCs explain 90.2% of the total variance vertically, so $m^* = 2$ will be used in the sequel.

The coordinates of the first two PCs at the leading Fourier frequencies are shown in Table 1: the absolute values, further, the real and imaginary parts of the coordinates. As for PC1, the leading frequency is 0.458 that indicates a period $2\pi/0.458 = 13.718$ which is about a biweekly period to which ISE has the largest contribution. As for PC2, the leading frequency is 2.866 that indicates a period $2\pi/2.866 = 2.193$ which is about a 2 days period to which BOVESPA has the largest contribution.

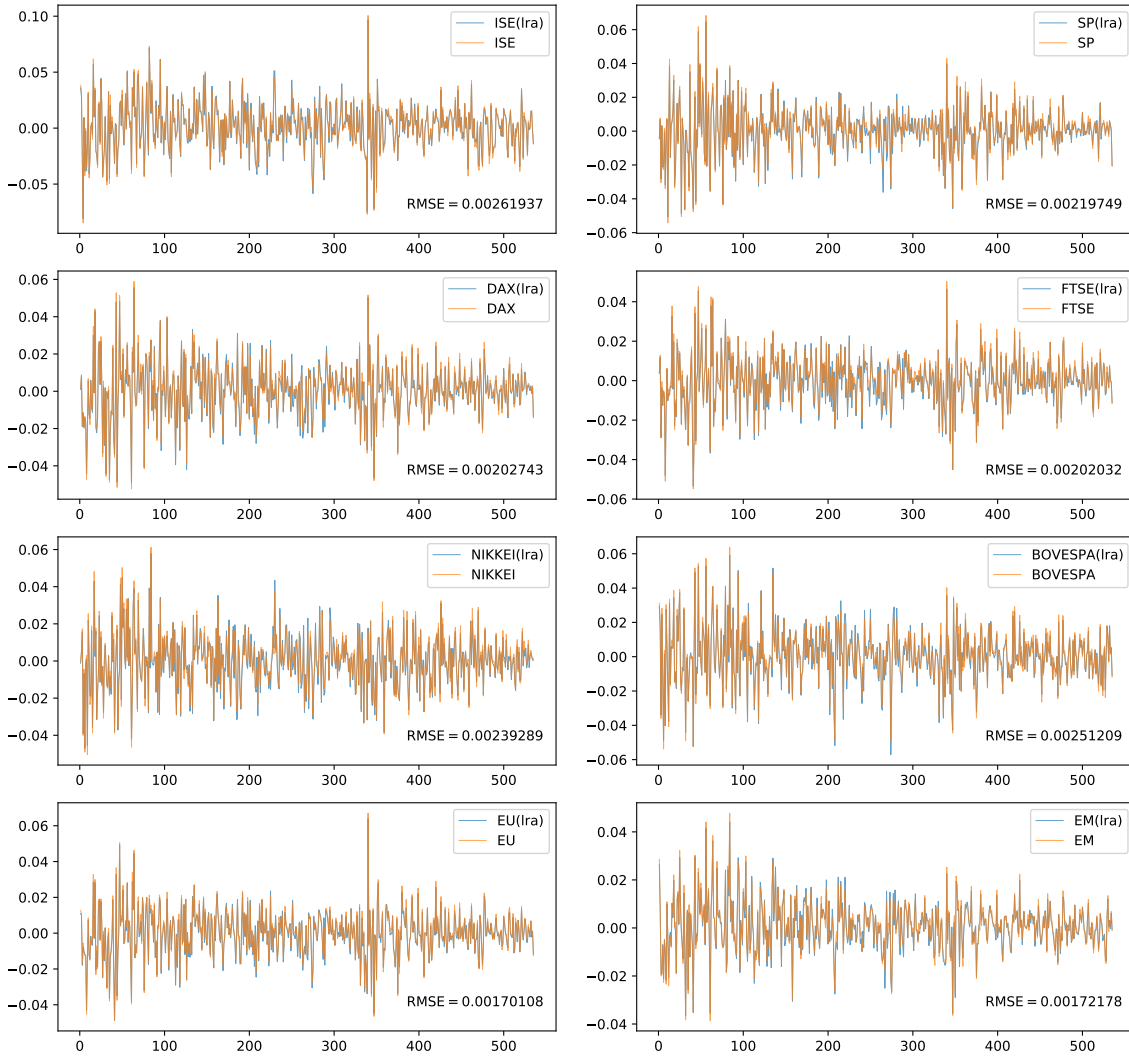


Figure 2: Rank 3 approximation of the variables (blue) on the original data (orange).

4.5 Block Cholesky decomposition in the time domain

It is well known (see e.g., [7]) that a d -dimensional regular process $\{\mathbf{X}_t\}$ with spectral density matrix \mathbf{f} of rank $r \leq d$ has the multidimensional Wold decomposition

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{B}_j \mathbf{Y}_{t-j},$$

where \mathbf{B}_j s are $d \times d$ matrices (like dynamic factor loadings), and $\{\mathbf{Y}_t\}$ is d -dimensional white noise, $\mathbb{E}\mathbf{Y}_t = \mathbf{0}$ and $\mathbb{E}\mathbf{Y}_t \mathbf{Y}_s^T = \delta_{ts} \boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}$ is $d \times d$ positive semidefinite matrix of rank r .

It is important that there is a one-to-one correspondence between \mathbf{f} (frequency domain) and the $\mathbf{B}(z)$, $\boldsymbol{\Sigma}$ pair (time domain), where $\mathbf{B}(z)$ is the transfer function

$$\mathbf{B}(z) = \sum_{j=0}^{\infty} \mathbf{B}_j z^j, \quad |z| \leq 1 \quad (10)$$

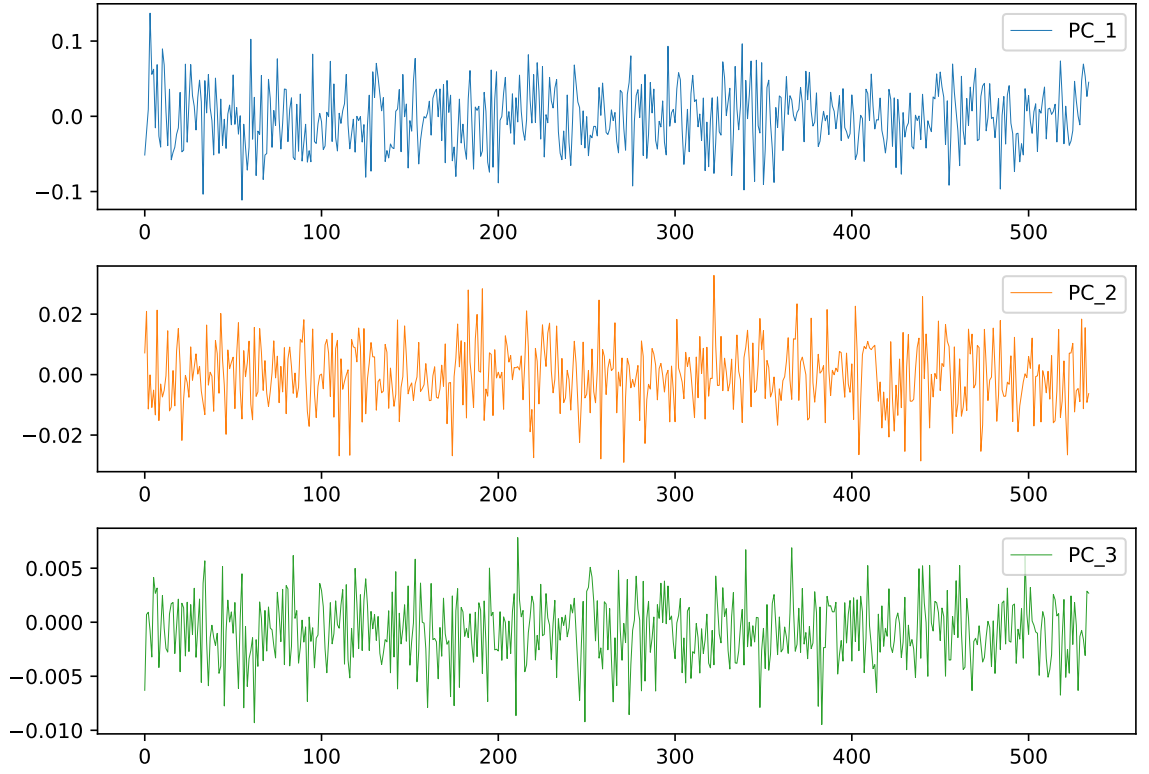


Figure 3: The 3 leading PCs of the stock exchange data in the time domain, colored according to the eigenvalues in Fig. 1.

with impulse responses \mathbf{B}_j s. This correspondence is given by

$$\mathbf{f}(z) = \frac{1}{2\pi} \mathbf{B}(z) \boldsymbol{\Sigma} \mathbf{B}^*(z), \quad |z| \leq 1.$$

So, via the spectral decomposition of \mathbf{f} and by performing the expansion (10) at the Fourier frequencies, we can estimate the transfer function.

In practice, only finitely many observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ are available. Then get the recursive equations

$$\mathbf{X}_k = \sum_{j=1}^{k-1} \mathbf{B}_{kj} \mathbf{Y}_j + \mathbf{Y}_k, \quad k = 1, 2, \dots, n. \quad (11)$$

If our process is stationary, the coefficient matrices are irrespective of the choice of the starting time, and \mathbf{Y}_j is the innovation, which is rather considered as the added value of observation j to the process than an error term. Its covariance matrix is $\mathbf{E}_j = \mathbb{E} \mathbf{Y}_j \mathbf{Y}_j^T$.

Multiplying the equations in (11) by \mathbf{X}_j^T from the right, and taking expectation, the solution for the matrices \mathbf{B}_{kj} and \mathbf{E}_j ($k = 1, \dots, n$; $j = 1, \dots, k-1$) can be obtained via the block Cholesky (LDL) decomposition:

$$\mathcal{C}_n = \mathbf{L}_n \mathbf{D}_n \mathbf{L}_n^T, \quad (12)$$

where \mathcal{C}_n is $nd \times nd$ positive definite block Toeplitz matrix of Section 2, \mathbf{D}_n is $nd \times nd$ block diagonal and contains the positive semidefinite prediction error matrices $\mathbf{E}_1, \dots, \mathbf{E}_n$ in its diagonal blocks, whereas \mathbf{L}_n is $nd \times nd$ lower triangular with blocks \mathbf{B}_{kj} s below its diagonal blocks which are $d \times d$ identities, so \mathbf{L}_n is

Table 1: The loadings inside the first two principal components at the frequency (on scale $[0, \pi]$) of the maximal eigenvalue

	PC1			PC2		
	abs. val.	real part	imag. part	abs. val.	real part	imag. part
max_freq_at	0.458			2.866		
ISE	0.498	0.498	0.000	0.529	0.529	-0.000
SP	0.302	0.301	0.019	0.393	-0.383	0.086
DAX	0.372	0.370	-0.040	0.255	-0.160	0.199
FTSE	0.297	0.297	-0.013	0.190	-0.124	0.144
NIKKEI	0.389	0.347	-0.177	0.195	0.130	0.146
BOVESPA	0.316	0.316	-0.016	0.601	-0.573	0.180
EU	0.312	0.311	-0.024	0.246	-0.174	0.174
EM	0.292	0.260	-0.134	0.068	0.049	0.047

non-singular:

$$L_n = \begin{bmatrix} I & O & \dots & O & O \\ B_{21} & I & \dots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ B_{n1} & B_{n2} & \dots & B_{n,n-1} & I \end{bmatrix}, \quad D_n = \begin{bmatrix} E_1 & O & \dots & O & O \\ O & E_2 & \dots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \dots & O & E_n \end{bmatrix}.$$

To find the block Cholesky decomposition of (12), the following recursion is at our disposal: for $j = 1, \dots, n$

$$E_j := C(0) - \sum_{m=1}^{j-1} B_{jm} E_m B_{jm}^T, \quad j = 1, \dots, n \quad (13)$$

and for $i = j + 1, \dots, n$

$$B_{ij} := \left(C(i-j) - \sum_{k=1}^{j-1} B_{ik} E_k B_{ik}^T \right) E_j^-, \quad (14)$$

where we take generalized inverse $^-$ if $r < d$.

Note that Equation (12) implies the following:

$$|C_n| = |D_n| = \prod_{j=1}^n |E_j|.$$

Here E_n is the error covariance matrix of the prediction of \mathbf{X}_n based on its $(n-1)$ -length long past. In the stationary case, if we predict based on the n -length long past, then we project on a richer subspace, therefore the prediction errors of the linear combinations of the coordinates of \mathbf{X}_n are decreased (better to say, not increased). Consequently, the ranks of the error covariance matrices E_n s are also decreased (not increased) as $n \rightarrow \infty$.

If the prediction is based on the infinite past, then with $n \rightarrow \infty$ this procedure (which is a nested one) extends to the multidimensional Wold decomposition. The point is that here $t = n$, and at time t we look back only at the $t-1$ length long past. Also, if $n \rightarrow \infty$, the matrix L_n better and better approaches a Toeplitz one, and the matrices E_1, \dots, E_n are closer and closer to Σ , the covariance matrix of the innovation process. Consequently, $B_{nj} \rightarrow B_j$ as $n \rightarrow \infty$. Note that if the autocovariances are absolutely summable, then the spectral decomposition of C_n is asymptotically closer and closer to that of $C_n^{(s)}$ (as $n \rightarrow \infty$). Latter one is described by Equation (7) as $C_n^{(s)} = \mathbf{W} \Lambda^{(s)} \mathbf{W}^*$. Here \mathbf{W} is a Vandermonde matrix, and both \mathbf{W} and $\Lambda^{(s)}$ contain primitive n th roots in the succession described in Section 2. Therefore, it is not surprising that the Gram and block Cholesky decompositions of C_n show the asymptotic Toeplitz phenomenon, as the neighboring entries in parallel (to the diagonal) lines depend on neighboring primitive roots, which are close to each other as $n \rightarrow \infty$.

Going further, if we want to find a lower ($m < r$) rank approximation of the E_j s, then we find a system ξ_1, \dots, ξ_n by means of projections within the innovation subspaces, $\mathbb{E}\xi_j\xi_j^T = I_m$. Actually, $\xi_j \in \mathbb{R}^m$ is the principal component factor of Y_j obtained from the m -factor model

$$Y_j = A_j \xi_j + \varepsilon_j,$$

where the columns of $d \times m$ matrix A_j are $\sqrt{\mu_{j\ell}} \mathbf{u}_{j\ell}$ ($\ell = 1, \dots, m$) with the m largest eigenvalues and the corresponding eigenvectors of E_j ; the vector ε_j is the error comprised of both the idiosyncratic noise and the error term of the model, but it has a negligible L^2 -norm. Note that A_j of the decomposition $E_j = A_j A_j^T$ is far not unique, it can be post-multiplied with an $m \times m$ orthogonal matrix. With this,

$$X_k \sim \sum_{j=1}^k B_{kj} A_j \xi_j, \quad k = 1, 2, \dots, n, \quad (15)$$

where $B_{kk} = I_k$. This approaches the following Wold decomposition of the d -dimensional process $\{X_t\}$ with a m -dimensional ($m \leq r \leq d$) innovation process $\{\xi_t\}$:

$$X_t \sim \sum_{j=0}^{\infty} B_j A \xi_{t-j},$$

where the $d \times m$ matrix A comes from the Gram decomposition $\Sigma = AA^T$ if the matrix Σ is of rank m ; otherwise, AA^T is just the best m -rank approximation of Σ .

Note that here we use $nd \times nd$ block matrices, but the procedure realized by equations (13) and (14) iterates only with the $d \times d$ blocks of it, so the computational complexity of this algorithm is $O\left(\binom{n}{2} d^3\right)$. However, this issue is rather the topic of Dynamic Factor Analysis, see [8].

5 Conclusions

Given a weakly stationary, multivariate time series with absolutely summable entries, we proved asymptotic relation between the eigenvalues of the block Toeplitz matrix of the first n autocovariances and the union of spectra of the spectral density matrices at the n Fourier frequencies, as $n \rightarrow \infty$. For the proof, eigenvalues of block circulant matrices are used, together with eigenvectors of both real and complex coordinates.

The theorem unveils important analogies between the time and frequency domain calculations. For instance, the complex principal components in the frequency domain are used for low-rank approximation of the process in the time domain; whereas, the block Cholesky decomposition of the block Toeplitz matrix gives rise to dimension reduction within the innovation subspaces, see Section 4.5. Above touching upon the computational benefits of our algorithm, the calculations are illustrated on a real-life financial time series.

Acknowledgement: The research reported in this paper and carried out at the Budapest University of Technology and Economics has been supported by the National Research Development and Innovation Fund based on the charter of bolster issued by the National Research Development and Innovation Office under the auspices of the Ministry for Innovation and Technology; also supported by the National Research, Development and Innovation Fund (TUDFO/51757/2019-ITM, Thematic Excellence Program). The research was funded by the EFOP-3.6.2-16-2017-00015-HU-MATHS-IN Hungarian Industrial and Innovation project, under the auspices of the István Széchenyi University, Győr.

We are also indebted to Katalin Friedl for valuable information about the eigenvalues of circulant matrices and permutation matrices.

Data Availability Statement: The third-party dataset analyzed during the current study is available in the UCI Machine Learning Repository, and was collected by the authors of [6]. The dataset is available in: Dua, D. and Graff, C. (2019). UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science, <https://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE>

References

- [1] P. J. Brockwell, R. A. Davis, and S. E. Fienberg. *Time series: Theory and methods*. Springer Science & Business Media, 1991.
- [2] D. R. Brillinger. *Time series: Data analysis and theory*, volume 36. SIAM, 1981.
- [3] G. J. Tee. Eigenvectors of block circulant and alternating circulant matrices. *New Zealand Journal of Mathematics*, 36(8):195–211, 2007.
- [4] B. Friedman. Eigenvalues of composite matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 57(1):37–49, 1961.
- [5] C. R. Rao. *Linear statistical inference and its applications*, volume 2. Wiley, 1973.
- [6] Oguz Akbilgic, Hamparsum Bozdogan, and M. Erdal Balaban. A novel Hybrid RBF Neural Networks model as a forecaster. *Statistics and Computing*, 24(3):365–375, May 2014.
- [7] M. Deistler and W. Scherrer. *Modelle der Zeitreihenanalyse*. Springer, 2018.
- [8] M. Bolla. Factor analysis, dynamic. *Wiley StatsRef: Statistics Reference Online*, pages 1–15, 2014.