

Asymptotic Theory of Principal Component Analysis for Time Series Data with Cautionary Comments

Xinyu Zhang¹ and Howell Tong²

¹*Center for Statistical Science, Department of Industrial Engineering, Tsinghua University, Beijing, China.*

²*School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China; Tsinghua University, Beijing, China; London School of Economics and Political Science, London, UK.*

Summary. Principal component analysis (PCA) is a most frequently used statistical tool in almost all branches of data science. However, like many other statistical tools, there is sometimes the risk of misuse or even abuse. In this paper, we highlight possible pitfalls in using the theoretical results of PCA based on the assumption of independent data when the data are time series. For the latter, we state with proof a central limit theorem of the eigenvalues and eigenvectors (loadings), give direct and bootstrap estimation of their asymptotic covariances, and assess their efficacy via simulation. Specifically, we pay attention to the proportion of variation, which decides the number of principal components (PCs), and the loadings, which help interpret the meaning of PCs. Our findings are that while the proportion of variation is quite robust to different dependence assumptions, the inference of PC loadings requires careful attention. We initiate and conclude our investigation with an empirical example on portfolio management, in which the PC loadings play a prominent role. It is given as a paradigm of correct usage of PCA for time series data.

Keywords: Bootstrap, Inference, Limiting distribution, PCA, Portfolio management, Time series.

1. Introduction

Principal component analysis (PCA) is probably one of the most widely used statistical tools in statistics and fields well beyond. Comprehensive summaries and numerous em-

pirical examples of PCA can be found in books such as [Flury \(1997\)](#), [Jolliffe \(2002\)](#) and [Tsay \(2005, 2014\)](#). Now, classical results for PCA are obtained under the assumption of *independent* (vectorial) data. However, applications to dependent data abound in which dependence is ignored. To cite but a few examples, we mention [Stone \(1947\)](#) in economics, [Craddock \(1965\)](#) and [Maryon \(1979\)](#) in meteorology, [Ahamad \(1967\)](#) for crime rates, [Feeney and Hester \(1967\)](#), [Tsay \(2005\)](#), [Wei \(2019\)](#) and [Keijsers and Dijk \(2019\)](#) in finance, and others. The issues that arise naturally are the defects, in both theory and practice, of this misuse, and a correct inference of PCA for time series. These are the issues that this paper aims to resolve.

For independent data, classical results about the limiting distribution of the principal components (PCs) can be found in [Anderson \(2003\)](#) and are widely known. For dependent data, a comprehensive and practically relevant central limit theorem for time series data is lacking in the literature. In this paper, we will state with proof a central limit theorem for stationary and ergodic multivariate linear time series by building on [Hannan \(1976\)](#). Besides the limiting distributions of estimated eigenvalues and estimated eigenvectors, we also give the limiting distribution of the proportion of variation to facilitate the decision of PC numbers. Then, we introduce two special cases or approximations of the theorem when the processes are Gaussian, which assume simpler dependence structure, one being the classical result for independent data. The efficacy of our theorem and the two special cases are assessed via simulation for Gaussian time series. It turns out that the proportion of variation is a statistic quite robust to different modes of dependence, in that they have only negligible effect. However, the situation with the loadings is markedly different. In the simulation, the efficacy of our theorem is verified, while the two special cases behave poorly for the loadings numerically and thus are not practically useful for time series data.

Inference based on the limiting distribution necessitates the estimation of the asymptotic covariances of eigenvalues and eigenvectors. For this, we give two methods: the direct and the bootstrap estimation. The former is recommended for Gaussian processes and some light-tail processes, while the latter is applicable to general processes.

In practice, PCA for time series is widely implemented but inference is often missing or inappropriate. To illustrate, we focus on a specific application field of PCA, namely

the stock portfolio management. Portfolios can be constructed based on the principal components and they are named “principal portfolios” (Partovi and Caputo, 2004). In Pasini (2017), it is shown that these principal portfolios can get enhanced returns and financial risk control. Further, they can be viewed as individual assets, based on which all the asset allocation strategies can be applied to reduce risk. See, e.g., Yang (2015). An important benefit is that in principal portfolios investors do not need to concern themselves with the co-movements among assets. The first principal portfolio is normally understood as the market component with roughly equal contributions of the underlying stocks. A number of following principal portfolios always represent synchronized fluctuations that only happen to a group of stocks.

The obtained principal portfolios are each a linear combination of all the stocks. However, although diversification can help reducing risks, the benefit will decrease as the number of stocks increases. Investors always prefer a portfolio that is sufficiently diverse, at the same time being as small as possible. Thus, implementing inference to the principal portfolio loadings is important since a much smaller portfolio could be obtained by discarding stocks with insignificant loadings. The sparser loadings will also lead to a more refined interpretation. In the real data example, we indicate how possible principal portfolios could be constructed by performing a proper inference on the loadings of the PCA. By so doing, we give a paradigm of correct procedures of implementing PCA on time series, and highlight the fact that without an appropriate inference, mis-interpretations in general, and wrong portfolio strategy in particular, can arise.

The paper is organised as follows. In Section 2, we give with proof a central limit theorem of the principal components for time series, and obtain the limiting distributions of eigenvalues, eigenvectors and the proportion of variation. To compare, in Section 3 we introduce two special cases of the theorem for Gaussian processes, one being the classical result for independent data. To estimate the asymptotic covariances of the eigenvalues and eigenvectors, two methods, the direct and the bootstrap estimation, are given in Section 4. In Section 5, we conduct simulations to assess the efficacy of the above theoretical results and highlight pitfalls of using classical results (based on the assumption of independent data) for time series. We also evaluate the numerical performance of the two estimation

methods. An empirical example of the stocks portfolio is given in Section 6 to show a paradigm of correct implementation of PCA for time series data. We conclude in Section 7. Details about the simulation are in Appendix A.

2. A central limit theorem

Let $\mathbf{X}(t)$ be a stationary and ergodic p -dimensional linear vector process generated as

$$\mathbf{X}(t) = \sum_{j=0}^{\infty} \mathbf{G}(j)\mathbf{e}(t-j), \quad t \in \{\dots, -1, 0, 1, \dots\}, \quad (1)$$

where $\mathbf{e}(t)$ is p -dimensional, and $E(\mathbf{e}(t)) = \mathbf{0}_p$, $E(\mathbf{e}(t_1)\mathbf{e}(t_2)^T) = \mathbf{K}\mathbf{1}(t_1 = t_2)$ with \mathbf{K} a $p \times p$ nonsingular matrix. Here $\mathbf{1}(\cdot)$ is the indicator function. The Kronecker delta $\delta_{t_1 t_2}$ is 1 when $t_1 = t_2$ and 0 otherwise. Here $\mathbf{G}(j)$'s are $p \times p$ matrices with $\sum_{j=0}^{\infty} \text{tr}\{\mathbf{G}(j)\mathbf{K}\mathbf{G}(j)^T\} < \infty$.

Let $\mathbf{\Gamma}(s) = \mathbf{\Gamma}_{\mathbf{X}}(s) = E(\mathbf{X}(t+s)\mathbf{X}(t)^T)$ for $s \in \{\dots, -1, 0, 1, \dots\}$. Specifically, let $\mathbf{\Gamma} \equiv \mathbf{\Gamma}(0)$ be the covariance matrix. Then, the spectral density matrix of $\mathbf{X}(t)$ at frequency ω is defined as

$$f(\omega) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \exp\{-is\omega\}\mathbf{\Gamma}(s),$$

where $i = (-1)^{1/2}$. Let the transfer function be $h(\omega) = \sum_{j=0}^{\infty} \mathbf{G}(j)e^{-ij\omega}$.

Suppose we have observations $\{\mathbf{X}(1), \dots, \mathbf{X}(n)\}$. Define the sample covariance matrix as $\mathbf{C}_n = (1/n) \sum_{t=1}^n (\mathbf{X}(t) - \bar{\mathbf{X}})(\mathbf{X}(t) - \bar{\mathbf{X}})^T$, where $\bar{\mathbf{X}}$ is the sample mean. Denote the (i, j) -th component of \mathbf{C}_n , $\mathbf{\Gamma}$, $h(\omega)$ and $f(\omega)$ by $\mathbf{C}_{n,ij}$, $\mathbf{\Gamma}_{ij}$, $h_{ij}(\omega)$ and $f_{ij}(\omega)$, respectively. Denote the conjugate of $f_{ij}(\omega)$ by $\overline{f_{ij}(\omega)}$.

Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$ and $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_p)$ be the eigenvalues and eigenvectors of $\mathbf{\Gamma}$, respectively. Assume all eigenvalues are positive and distinct. Specifically, $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$. Let $\mathbf{l} = (l_1, \dots, l_p)$ and $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_p)$ be the eigenvalues and eigenvectors of \mathbf{C}_n , respectively. Let $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ and $\mathbf{L} = \text{diag}(\mathbf{l})$. Then we can express $\mathbf{\Gamma} = \boldsymbol{\Phi}\mathbf{\Lambda}\boldsymbol{\Phi}^T$ and $\mathbf{C}_n = \mathbf{A}\mathbf{L}\mathbf{A}^T$.

When applying PCA, we pay attention to the number and loadings of PCs. The number of PCs are usually decided by the proportion of variation

$$\gamma_k = \sum_{i=1}^k \lambda_i / \sum_{i=1}^p \lambda_i = \sum_{i=1}^k \lambda_i / \text{tr}(\mathbf{\Gamma}), \quad r_k = \sum_{i=1}^k l_i / \sum_{i=1}^p l_i = \sum_{i=1}^k l_i / \text{tr}(\mathbf{C}_n), \quad (2)$$

for $k = 1, \dots, p$. Let $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)$ and $\mathbf{r} = (r_1, \dots, r_p)$. To derive the asymptotics, we introduce two assumptions.

ASSUMPTION 1. Let \mathcal{F}_t be the sub σ -algebra generated by $\{\mathbf{X}(n), n \leq t\}$. Denote the a -th component of $\mathbf{e}(t)$ as $e_a(t)$, $a = 1, \dots, p$. Suppose that

$$\begin{aligned} & \mathbb{E}\{e_a(t)|\mathcal{F}_{t-1}\}, \quad \mathbb{E}\{e_a(t)e_b(t)|\mathcal{F}_{t-1}\}, \\ & \mathbb{E}\{e_a(t)e_b(t)e_c(t)|\mathcal{F}_{t-1}\}, \quad \mathbb{E}\{e_a(t)e_b(t)e_c(t)e_d(t)|\mathcal{F}_{t-1}\}, \end{aligned} \quad (3)$$

are constants for all $a, b, c, d = 1, \dots, p$.

ASSUMPTION 2. Suppose $f_{kk}(\omega)$, $k = 1, 2, \dots, p$, are all square integrable.

The following lemma is from Hannan (1976), after correcting some (possibly typographical) errors in the second term of his equation (3).

LEMMA 1. Suppose $\mathbf{X}(t)$ is stationary, ergodic, and generated by (1). Then, under Assumptions 1 and 2, for $i, j, k, l = 1, \dots, p$, $\sqrt{n}\{\mathbf{C}_{n,ij} - \boldsymbol{\Gamma}_{ij}\}$ has a joint asymptotic normal distribution with zero mean, and the asymptotic covariance between $\sqrt{n}\{\mathbf{C}_{n,ij} - \boldsymbol{\Gamma}_{ij}\}$ and $\sqrt{n}\{\mathbf{C}_{n,kl} - \boldsymbol{\Gamma}_{kl}\}$ is

$$\begin{aligned} & 2\pi \int_{-\pi}^{\pi} \left\{ f_{ik}(\omega) \overline{f_{jl}(\omega)} + f_{il}(\omega) \overline{f_{jk}(\omega)} \right\} d\omega \\ & + (4\pi^2)^{-1} \sum_{a,b,c,d=1}^p \kappa_{abcd} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h_{ia}(\omega_1) \overline{h_{jb}(\omega_1)} \overline{h_{kc}(\omega_2)} h_{ld}(\omega_2) d\omega_1 d\omega_2, \end{aligned} \quad (4)$$

where κ_{abcd} is the joint fourth cumulant of $e_a(t)$, $e_b(t)$, $e_c(t)$ and $e_d(t)$. Specifically,

$$\begin{aligned} \kappa_{abcd} = & \mathbb{E}\{e_a(t)e_b(t)e_c(t)e_d(t)\} - \mathbb{E}\{e_a(t)e_b(t)\}\mathbb{E}\{e_c(t)e_d(t)\} \\ & - \mathbb{E}\{e_a(t)e_c(t)\}\mathbb{E}\{e_b(t)e_d(t)\} - \mathbb{E}\{e_a(t)e_d(t)\}\mathbb{E}\{e_b(t)e_c(t)\}. \end{aligned}$$

Let $\mathbf{Y}(t) = \boldsymbol{\Phi}^T \mathbf{X}(t)$ be the principal components of $\mathbf{X}(t)$, and $q(\omega) = \boldsymbol{\Phi}^T h(\omega)$ be its transfer function. Then, the spectral density matrix of $\mathbf{Y}(t)$ is

$$g(\omega) = \boldsymbol{\Phi}^T f(\omega) \boldsymbol{\Phi} = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \exp\{-is\omega\} \boldsymbol{\Gamma}_{\mathbf{Y}}(s),$$

with $\mathbf{\Gamma}_{\mathbf{Y}}(s) = \mathbb{E}(\mathbf{Y}(t+s)\mathbf{Y}(t)^T)$ for $s \in \{\dots, -1, 0, 1, \dots\}$. Although $\mathbf{\Gamma}_{\mathbf{Y}}(s)$ is diagonal at $s = 0$, it is not so at other s . Thus, different from the PCs obtained for i.i.d. data, which is free from spatial dependence, there is temporal spatial dependence in $\mathbf{Y}(t)$. Applying Lemma 1 to $\mathbf{Y}(t)$, Corollary 1 follows.

COROLLARY 1. *Let $\mathbf{T}_n = \mathbf{\Phi}^T \mathbf{C}_n \mathbf{\Phi}$ and $\mathbf{U} = \sqrt{n}(\mathbf{T}_n - \mathbf{\Lambda})$. Denote the (i, j) -th component of \mathbf{U} by u_{ij} . Then, $\{u_{ij}\}$ has a joint asymptotic normal distribution whose mean is zero and whose asymptotic covariance (denoted as \mathcal{C}) between $\{u_{ij}\}$ and $\{u_{kl}\}$ is*

$$\begin{aligned} \mathcal{C}(u_{ij}, u_{kl}) = & 2\pi \int_{-\pi}^{\pi} \left\{ g_{ik}(\omega) \overline{g_{jl}(\omega)} + g_{il}(\omega) \overline{g_{jk}(\omega)} \right\} d\omega \\ & + (4\pi^2)^{-1} \sum_{a,b,c,d=1}^p \kappa_{abcd} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} q_{ia}(\omega_1) \overline{q_{jb}(\omega_1) q_{kc}(\omega_2)} q_{ld}(\omega_2) d\omega_1 d\omega_2. \end{aligned} \quad (5)$$

Now, we state the main result, a central limit theorem for the eigenvalues, the eigenvectors and the proportion of variation as follows.

THEOREM 1. *Let $\mathbf{X}(t)$ satisfy the assumptions of Lemma 1 and suppose $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$. Define $\mathbf{D} = \sqrt{n}(\mathbf{L} - \mathbf{\Lambda})$, $\mathbf{H} = \sqrt{n}(\mathbf{A} - \mathbf{\Phi})$ and $\mathbf{V} = \mathbf{\Phi}^T \mathbf{H}$. Then, for the elements of the matrix $\mathbf{D} = \text{diag}(d_i)$ and $\mathbf{V} = \{v_{ij}\}$ for $i, j = 1, \dots, p$, the following results hold*

$$d_i = u_{ii} + o(1), \quad v_{ii} = o(1), \quad \text{and} \quad v_{ij} = u_{ij}/(\lambda_j - \lambda_i) + o(1), \quad i \neq j, \quad (6)$$

a.s. as $n \rightarrow \infty$. Then, the limiting distributions of \mathbf{D} and \mathbf{H} are obtained since the asymptotic distribution of $\{u_{ij}\}$ is given in Corollary 1.

Let \implies denote weak convergence. We show the asymptotics of the proportion of variation and loadings of PCs specifically.

(1) *The eigenvalues \mathbf{l} has the limiting distribution $\sqrt{n}(\mathbf{l} - \mathbf{\lambda}) \implies \mathcal{N}(0, \mathbf{B})$, where $\mathbf{B} = \{b_{ij}\}_{i,j=1,\dots,p}$ with $b_{ij} = \mathcal{C}(u_{ii}, u_{jj})$.*

For $k = 1, \dots, p$, the proportion of variation r_k has the limiting distribution

$$\sqrt{n}(r_k - \gamma_k) \implies \mathcal{N}(0, \eta_k^2), \quad \eta_k^2 = \sum_{i=1}^p \sum_{j=1}^p \mathcal{C}(u_{ii}, u_{jj}) \beta_i(k) \beta_j(k), \quad (7)$$

with $\beta_i(k) = (\mathbf{1}(i \leq k) - \gamma_k)/\text{tr}(\mathbf{\Gamma})$.

(2) For $k = 1, \dots, p$, the PC loadings \mathbf{a}_k has the limiting distribution

$$\sqrt{n}(\mathbf{a}_k - \boldsymbol{\phi}_k) \implies \mathcal{N}(0, \boldsymbol{\Sigma}_k), \quad \boldsymbol{\Sigma}_k = \sum_{i=1, i \neq k}^p \sum_{j=1, j \neq k}^p \frac{\mathcal{C}(u_{ik}, u_{jk})}{(\lambda_k - \lambda_i)(\lambda_k - \lambda_j)} \boldsymbol{\phi}_i \boldsymbol{\phi}_j^T. \quad (8)$$

PROOF. Let $\mathbf{Z} = \boldsymbol{\Phi}^T \mathbf{A}$, then $\mathbf{T}_n = \mathbf{Z} \mathbf{L} \mathbf{Z}^T$, and $\mathbf{V} = \sqrt{n}(\mathbf{Z} - \mathbf{I})$. By the same argument as [Anderson \(2003\)](#) (Page 546), substituting $\mathbf{T}_n = \boldsymbol{\Lambda} + (1/\sqrt{n})\mathbf{U}$, $\mathbf{Z} = \mathbf{I} + (1/\sqrt{n})\mathbf{V}$ and $\mathbf{L} = \boldsymbol{\Lambda} + (1/\sqrt{n})\mathbf{D}$ in the equations $\mathbf{T}_n = \mathbf{Z} \mathbf{L} \mathbf{Z}^T$ and $\mathbf{I} = \mathbf{Z} \mathbf{Z}^T$, the result in (6) follows. Using the fact that $\mathbf{H} = \boldsymbol{\Phi} \mathbf{V}$ and [Corollary 1](#), the limiting distributions of \mathbf{D} and \mathbf{H} are obtained. Using the delta method, the limiting distribution of r_k is obtained by noting the fact that it is a differentiable function of \mathbf{l} . Thus, the proof is complete. \square

For Gaussian processes, κ_{abcd} is zero, and (5) is simplified to

$$\mathcal{C}(u_{ij}, u_{kl}) = 2\pi \int_{-\pi}^{\pi} \left\{ g_{ik}(\omega) \overline{g_{jl}(\omega)} + g_{il}(\omega) \overline{g_{jk}(\omega)} \right\} d\omega. \quad (9)$$

COROLLARY 2. Let $\mathbf{X}(t)$ be a Gaussian process that satisfies the assumptions of [Theorem 1](#). Then, the following asymptotic results hold for $k = 1, 2, \dots, p$.

(1) Jointly, \mathbf{l} and \mathbf{a}_k are asymptotically normally distributed and asymptotically unbiased.

(2) For \mathbf{l} , $\sqrt{n}(\mathbf{l} - \boldsymbol{\lambda}) \implies \mathcal{N}(0, \mathbf{B})$, where $b_{ij} = 4\pi \int_{-\pi}^{\pi} |g_{ij}(\omega)|^2 d\omega$. Thus, (7) is now

$$\sqrt{n}(r_k - \gamma_k) \implies \mathcal{N}(0, \eta_k^2), \quad \text{where } \eta_k^2 = \sum_{i=1}^p \sum_{j=1}^p 4\pi \int_{-\pi}^{\pi} |g_{ij}(\omega)|^2 d\omega \beta_i(k) \beta_j(k). \quad (10)$$

(3) For $\sqrt{n}(\mathbf{a}_k - \boldsymbol{\phi}_k) \implies \mathcal{N}(0, \boldsymbol{\Sigma}_k)$, (8) is now

$$\boldsymbol{\Sigma}_k = \sum_{i=1, i \neq k}^p \sum_{j=1, j \neq k}^p \frac{2\pi \int_{-\pi}^{\pi} \left\{ g_{ij}(\omega) \overline{g_{kk}(\omega)} + g_{ik}(\omega) \overline{g_{kj}(\omega)} \right\} d\omega}{(\lambda_k - \lambda_i)(\lambda_k - \lambda_j)} \boldsymbol{\phi}_i \boldsymbol{\phi}_j^T. \quad (11)$$

3. Special cases for Gaussian processes

As we have said, generally speaking, there is temporal spatial dependence in \mathbf{Y}_t . However, in some special cases, \mathbf{Y}_t has simpler dependence structure, thus leading to neater results. We discuss two such cases for Gaussian processes here.

- (a) All $\boldsymbol{\Gamma}_{\mathbf{Y}}(s)$ are diagonal, which means \mathbf{Y}_t are independently evolving time series, i.e., there is only temporal dependence in \mathbf{Y}_t . It follows that $g(\omega)$ is diagonal.

- (b) $\Gamma_{\mathbf{Y}}(0)$ is diagonal and all other $\Gamma_{\mathbf{Y}}(s)$ are zero, which means \mathbf{Y}_t are independent vectors, i.e., there is no temporal or spatial dependence in \mathbf{Y}_t .

The first case has been studied by [Taniguchi and Krishnaiah \(1987\)](#), although it is difficult to envisage a real situation where $\Gamma_{\mathbf{Y}}(s)$ is diagonal for *all* s . Perhaps, this case could be viewed as an approximation of Theorem 1, i.e., substituting $g(\omega)$ by its diagonalized version. The second case covers the classical PCA for independent data \mathbf{X}_t . It is developed for Gaussian data and has been extensively studied. See, e.g., [Anderson \(2003\)](#). To be consistent, we state the asymptotic results for Gaussian processes under (a) and (b) in Corollary 3 and 4, respectively.

COROLLARY 3. *Let $\mathbf{X}(t)$ be a Gaussian process that satisfies the assumptions of Theorem 1, and suppose all $\Gamma_{\mathbf{Y}}(s)$ are diagonal. Then, in (5), $\mathcal{C}(u_{ij}, u_{kl})$ is zero except when $i = j = k = l$, $i = k \neq j = l$ or $i = l \neq k = j$,*

$$\mathcal{C}(u_{ii}, u_{ii}) = 4\pi \int_{-\pi}^{\pi} g_{ii}^2(\omega) d\omega, \quad \mathcal{C}(u_{ij}, u_{ij}) = \mathcal{C}(u_{ij}, u_{ji}) = 2\pi \int_{-\pi}^{\pi} g_{ii}(\omega) g_{jj}(\omega) d\omega, \quad i \neq j.$$

which means the elements of \mathbf{U} are asymptotically independent. Thus, the following asymptotic results hold for $k = 1, 2, \dots, p$.

- (1) *Jointly, $\mathbf{1}$ and \mathbf{a}_k are asymptotically normally distributed and asymptotically unbiased.*
 (2) *For $\mathbf{1}$, $\sqrt{n}(\mathbf{1} - \boldsymbol{\lambda}) \implies \mathcal{N}(0, \mathbf{B})$, where $b_{ij} = 4\pi \int_{-\pi}^{\pi} g_{ii}^2(\omega) d\omega \mathbf{1}(i = j)$. Thus, (7) is now*

$$\sqrt{n}(r_k - \gamma_k) \implies \mathcal{N}(0, \eta_k^2), \quad \text{where } \eta_k^2 = \sum_{i=1}^p 4\pi \int_{-\pi}^{\pi} g_{ii}^2(\omega) d\omega \beta_i^2(k).$$

- (3) *For $\sqrt{n}(\mathbf{a}_k - \boldsymbol{\phi}_k) \implies \mathcal{N}(0, \boldsymbol{\Sigma}_k)$, (8) is now*

$$\boldsymbol{\Sigma}_k = \sum_{i=1, i \neq k}^p \frac{2\pi \int_{-\pi}^{\pi} g_{kk}(\omega) g_{ii}(\omega) d\omega}{(\lambda_k - \lambda_i)^2} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T.$$

- (4) *All of the l_k are asymptotically independent of all of the \mathbf{a}_k .*

COROLLARY 4. *Let $\mathbf{X}(t)$ be a Gaussian process that satisfies the assumptions of Theorem 1. Suppose $f(\omega) = \Gamma/(2\pi)$, which means $\mathbf{X}(t)$ is i.i.d. and $\mathcal{N}(0, \Gamma)$, and it follows that*

Table 1. Comparisons among Corollary 2, 3 and 4 for Gaussian processes. A ✓ means yes and a ✗ means no.

Corollary	2	3	4
Dependence in Y_t			
Temporal	✓	✓	✗
Spatial	✓	✗	✗
Asymptotic properties			
Joint Gaussianity of \mathbf{l} and \mathbf{a}_k	✓	✓	✓
Unbiasedness of \mathbf{l} and \mathbf{a}_k	✓	✓	✓
Independence among l_k	✗	✓	✓
Independence between all l_k and \mathbf{a}_k	✗	✓	✓

$g(\omega) = \mathbf{\Lambda}/(2\pi)$. Then, in (5), $\mathcal{C}(u_{ij}, u_{kl})$ is zero except when $i = j = k = l$, $i = k \neq j = l$ or $i = l \neq k = j$,

$$\mathcal{C}(u_{ii}, u_{ii}) = 2\lambda_i^2, \quad \mathcal{C}(u_{ij}, u_{ij}) = \mathcal{C}(u_{ij}, u_{ji}) = \lambda_i\lambda_j, \quad i \neq j,$$

which means the elements of \mathbf{U} are asymptotically independent. Thus, the following asymptotic results hold for $k = 1, 2, \dots, p$.

(1) Jointly, \mathbf{l} and \mathbf{a}_k are asymptotically normally distributed and asymptotically unbiased.

(2) For \mathbf{l} , $\sqrt{n}(\mathbf{l} - \boldsymbol{\lambda}) \implies \mathcal{N}(0, \mathbf{B})$, where $b_{ij} = 2\lambda_i^2 \mathbf{1}(i = j)$. Thus, (7) is now

$$\sqrt{n}(r_k - \gamma_k) \implies \mathcal{N}(0, \eta_k^2), \quad \text{where } \eta_k^2 = \sum_{i=1}^p 2\lambda_i^2 \beta_i^2(k).$$

(3) For $\sqrt{n}(\mathbf{a}_k - \boldsymbol{\phi}_k) \implies \mathcal{N}(0, \boldsymbol{\Sigma}_k)$, (8) is now

$$\boldsymbol{\Sigma}_k = \sum_{i=1, i \neq k}^p \frac{\lambda_k \lambda_i}{(\lambda_k - \lambda_i)^2} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T.$$

(4) All of the l_k are asymptotically independent of all of the \mathbf{a}_k .

We compare the results of Corollary 2, 3 and 4 for Gaussian processes in Table 1. Intuitively, a simpler dependence structure would bring benefits such as asymptotic independence or simpler forms of the asymptotic covariance. It should be noted that Corollaries 2,

3 and 4 study the same estimate \mathbf{I} , r_k and \mathbf{A} , but give different asymptotics. An essential question is how these different dependence structures affect the inference of the number and loadings of PCs in practice. Especially, we want to know the consequences of making inference under the assumption of independent observations (Corollary 4) when the real data is time series (Corollary 2). This situation is possibly the most common misuse of PCA for time series. Thus, we will discuss the estimation of the asymptotic covariance of eigenvalues and eigenvectors in the next Section, which is essential for the inference of the number and loadings of PCs.

4. Estimation of the asymptotic covariance of eigenvalues and eigenvectors

To conduct inference on the eigenvalues, eigenvectors and proportion of variation, an intuitively plausible approach is to produce directly consistent estimates of \mathbf{B} , Σ_k and η_k^2 from their analytical forms in (7) and (8). The main challenge lies in the second term of $\mathcal{C}(u_{ij}, u_{kl})$ in (5). In principle, this can be handled by the estimation of the integral of the fourth-order cumulant spectra as similar challenges often appear in the asymptotic covariances of quantities in time series analysis, such as spectral mean statistics (Shao, 2009), whittle estimation (Giraitis and Robinson, 2001) and so on. Now, this was the approach adopted by Taniguchi, Puri and Kondo (1996). However, their proposed estimator turns out to be computationally complex: its consistency requires the existence of the eighth order moment and the procedure involves a choice of a smoothing parameter with no theoretical guidance. As a result, in the literature, researchers have tended to avoid direct estimation of this term and preferred more appealing alternatives, such as the self-normalization approach (Shao, 2009) and the bootstrap method (Meyer, Paparoditis and Kreiss, 2020). Thus, we will only implement the direct estimation when κ_{abcd} is negligible, e.g., Gaussian processes, under which case Corollary 2 can be applied. Simulation results in the next section will show that this method has some scope of application for some light-tailed data.

Our experience suggests that a data-dependent bootstrap method can alleviate the troublesome aspects mentioned earlier and can be applied to general processes beyond the Gaussian.

4.1. Direct estimation

For a given Gaussian process, to estimate the asymptotic covariances \mathbf{B} , Σ_k and η_k^2 in Corollary 2, we need to estimate Φ , Λ and $g(\omega)$. For Φ and Λ , natural estimators are \mathbf{A} and \mathbf{L} . For the spectral density matrix $f(\omega)$, define the raw periodogram $\tilde{f}(\omega)$ and the smoothed periodogram $\hat{f}(\omega)$ as

$$\tilde{f}(\omega) = (2\pi)^{-1} \sum_{s=-(n-1)}^{n-1} \exp\{-is\omega\} \hat{\Gamma}(s), \quad \hat{f}(\omega) = \sum_{m=-M}^M W_M(m) \tilde{f}(\omega + m/n),$$

where $\hat{\Gamma}(s) = (1/n) \sum_{t=1}^n (\mathbf{X}(t+s) - \bar{\mathbf{X}})(\mathbf{X}(t) - \bar{\mathbf{X}})^T$, M is the bandwidth, and $W_M(m)$ is the spectral window function. For example, $W_M(m) = \mathbf{1}(|m| \leq M)/(2M+1)$ is the Daniell window. It is known that $\hat{f}(\omega)$ is a consistent estimator. Thus, the estimator of $g(\omega)$ is $\hat{g}(\omega) = \mathbf{A}^T \hat{f}(\omega) \mathbf{A}$. Then, we obtain the following estimators of \mathbf{B} , Σ_k and η_k^2 ,

$$\begin{aligned} \hat{b}_{ij} &= 4\pi \int_{-\pi}^{\pi} |\hat{g}_{ij}(\omega)|^2, \quad \hat{\eta}_k^2 = \sum_{i=1}^p \sum_{j=1}^p 4\pi \int_{-\pi}^{\pi} |\hat{g}_{ij}(\omega)|^2 d\omega \hat{\beta}_i(k) \hat{\beta}_j(k), \\ \hat{\Sigma}_k &= \sum_{i=1, i \neq k}^p \sum_{j=1, j \neq k}^p \frac{2\pi \int_{-\pi}^{\pi} \left\{ \hat{g}_{ij}(\omega) \overline{\hat{g}_{kk}(\omega)} + \hat{g}_{ik}(\omega) \overline{\hat{g}_{kj}(\omega)} \right\} d\omega}{(l_k - l_i)(l_k - l_j)} \mathbf{a}_i \mathbf{a}_j^T, \end{aligned}$$

with $\hat{\beta}_i(k) = (\mathbf{1}(i \leq k) - r_k) / \sum_{j=1}^p l_j$.

Similar procedures can be applied to obtain the direct estimation of the asymptotic covariances \mathbf{B} , Σ_k and η_k^2 in Corollary 3 and 4.

4.2. Bootstrap method

Our second method is to resample and derive the bootstrap covariance estimator, avoiding the difficulty and complexity of estimating the second term in (5). It is applicable to general processes, especially non-Gaussian and heavy-tailed processes. Various bootstrap methods for dependent data have been explored by researchers, including block bootstrap, sieve bootstrap, bootstrap in frequency domain, and others. Comprehensive reviews of bootstrap for dependent data can be found in Bühlmann (2002), Lahiri (2003) and Politis (2003).

Our experiences with the surrogate data method, the time frequency toggle method and the moving block bootstrap have narrowed the choice to the last method. The first two methods are in the frequency domain. The idea of the surrogate data method from

[Theiler et al. \(1992\)](#) is to bootstrap the phase of the Fourier coefficients but keep their magnitude unchanged. This method has limited validity because every surrogated sample has exactly the same periodogram (and mean) as the original series. For this reason, this method fails for our problem since it will generate exactly the same eigenvectors.

The time frequency toggle method was proposed by [Kirch and Politis \(2011\)](#), whose basic idea is to bootstrap the Fourier coefficients of the observed time series, and then to back-transform them to obtain a bootstrap sample in the time domain. It is more general than the surrogate data method in that it can capture the distribution of statistics based on the periodogram, and thus can work for our problem. However, our simulation results show that its performance for heavy-tailed data is not good. The reason may lie in the fact that the sample paths of this method are asymptotically Gaussian, due to the Fourier coefficients being asymptotically so. This seems inevitable for almost all methods using discrete Fourier transforms.

The moving block bootstrap (MBB) originated in [Hall \(1985\)](#) and [Carlstein \(1986\)](#), and was further developed in [Künsch \(1989\)](#), [Liu and Singh \(1992\)](#) and others for stationary observations. It resamples blocks of (consecutive) observations at a time. As a result, the dependence structure of the original observations is preserved within each block. To explain briefly, let us consider an observed univariate time series $X(1), \dots, X(n)$, for a parameter θ which is a functional of the j -dimensional marginal distribution of the time series, we consider the j -dimensional vectors

$$\mathbf{S}(t) = (X(t), \dots, X(t + j - 1)), \quad t = 1, \dots, n - j + 1.$$

For fixed block size ρ , define the blocks in terms of $\mathbf{S}(t)$ as

$$\Xi_t = (\mathbf{S}(t), \dots, \mathbf{S}(t + \rho - 1)), \quad t = 1, \dots, n + 2 - \rho - j.$$

For $\tau \geq 1$, select τ blocks randomly from the collection $\{\Xi_t, t = 1, \dots, n + 2 - \rho - j\}$, and align them to generate the MBB observations $\{\mathbf{S}^*(1), \dots, \mathbf{S}^*(\rho); \mathbf{S}^*(\rho + 1), \dots, \mathbf{S}^*(2\rho); \dots, \mathbf{S}^*(\tau\rho)\}$. And τ should be the smallest integer such that $\tau\rho \geq n$. Then, the estimator $\hat{\theta}$ is derived based on the MBB observations.

It can be seen that the above procedure can be easily modified for multivariate time series, so it can be applied to our PCA problem. However, the eigenvectors depend on

the entire distribution of the process, so the question of decision of j arises. An ad-hoc solution is to work with the naive block bootstrap (using $j = 1$). This will inevitably result in some efficiency loss compared to the ordinary block bootstrap, but the simulation in the next section suggests that the numerical performance is good and acceptable. So, we will use MBB for our problem, while also leaving the door open for other bootstrap methods.

5. Simulation

In this section, we will assess, via simulation, the asymptotics and estimation methods in above sections. Instead of focusing on the eigenvalues, eigenvectors and the proportion of variation ($\boldsymbol{\lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\gamma}$), our major interest is here in the asymptotic covariances of their estimates (\mathbf{l} , \mathbf{A} and \mathbf{r}), namely \mathbf{B} , $\boldsymbol{\Sigma}_k$ and η_k^2 , with different formulas given by Corollaries 2, 3 and 4 for Gaussian processes. In Section 5.1, we will check the efficacy of these formulas. The asymptotic distribution of Corollary 2 is denoted by AD. Results based on Corollaries 3 and 4 are abbreviated as DAG and IND to indicate the case with diagonalized spectral density matrix and the case of independent observations, respectively. Conducting PCA to simulated data and repeating over replications, we can get their empirical distribution, which is denoted as ED. By comparing the formula of \mathbf{B} , $\boldsymbol{\Sigma}_k$ and η_k^2 given by AD, DAG and IND respectively with the empirical covariance of $\sqrt{n}(\mathbf{l} - \boldsymbol{\lambda})$, $\sqrt{n}(\mathbf{a}_k - \boldsymbol{\phi}_k)$ and $\sqrt{n}(r_k - \gamma_k)$ given by ED, we can assess their performance.

In Section 5.2, we assess the two estimation methods, i.e. the direct method (DE) and the bootstrap method (BE) introduced in Section 4. For each simulation replication, we estimate the standard deviation of eigenvectors using the two estimation methods and averaging over replications. We compare them with the empirical standard deviations (ED). Here we handle not only Gaussian processes, but also non-Gaussian ones, including some with outliers, or skewed Gaussianity, or heavy-tail.

In the following simulation, we take finite-order Vector Autoregressive (VAR) and Vector Moving Average models (VMA) as examples, which are respectively defined as

$$\mathbf{X}(t) = \mathbf{e}(t) + \sum_{j=1}^J \mathbf{F}(j)\mathbf{X}(t-j), \quad \mathbf{X}(t) = \mathbf{e}(t) + \sum_{j=1}^J \mathbf{G}(j)\mathbf{e}(t-j),$$

where $\mathbf{e}(t)$ is a white noise series and J is the order.

5.1. Efficacy of the asymptotics for Gaussian processes

Let the length of time series be $n = 5000$, the dimension be $p = 5$, and the replication time be $N = 2000$. Suppose $\mathbf{e}(t)$ is from $\mathcal{N}(0_5, 10\mathbf{I}_5)$, with \mathbf{I}_5 being the identity matrix of dimension 5. Consider the following Data Generating Processes (DGPs):

- **DGP 1:** VAR(1) process with multivariate Gaussian noise.
- **DGP 2:** VMA(1) process with multivariate Gaussian noise.

Elements of $\mathbf{F}(j)$ and $\mathbf{G}(j)$ are obtained by first drawing random numbers from the uniform distribution and then transformed in order to ensure the stationarity of the VAR models and invertibility of the VMA models. And they are shown in Appendix A.1.

For the eigenvalues we will check their asymptotic covariance matrix $\text{Cov}(l_k, l_{k'}) \equiv b_{kk'}/n$ to show the asymptotic dependence among them. For the eigenvectors and the proportion of variation, we will check the asymptotic standard deviation of each element, $\sigma(a_{kk'}) \equiv \Sigma_{k,k'k'}/\sqrt{n}$ and $\sigma(r_k) \equiv \eta_k/\sqrt{n}$, for the sake of inference. We check $\sigma(r_k)$ for $k = 1, \dots, p-1$ since r_p is always 1 and thus $\sigma(r_p)$ is always 0. Note that they all have an empirical entry (ED), and three asymptotic entries (AD, DAG and IND), which will be marked in the subscript.

For $\sigma(r_k)$, we compute the difference (Δ) of AD with respect to ED in percentage as

$$\Delta_{\text{AD}}(r_k) = \sigma_{\text{AD}}(r_k) - \sigma_{\text{ED}}(r_k). \quad (12)$$

For $\sigma(a_{kk'})$, to avoid the influence of scale, we compute the difference ratio (Δ^*) of AD with respect to ED in percentage as

$$\Delta_{\text{AD}}^*(a_{kk'}) = \sigma_{\text{AD}}(a_{kk'})/\sigma_{\text{ED}}(a_{kk'}) - 1. \quad (13)$$

Similarly, we can compute $\Delta_{\text{DAG}}(r_k)$, $\Delta_{\text{IND}}(r_k)$, $\Delta_{\text{DAG}}^*(a_{kk'})$ and $\Delta_{\text{IND}}^*(a_{kk'})$.

For each model, the vectorized $\text{Cov}(l_k, l_{k'})$ and $\Delta^*(a_{kk'})$ for $k, k' = 1, \dots, 5$ are each a 25-dimensional vector. See their plots in Fig.1 for DGP 1 and 2 respectively. For the vectorized $\text{Cov}(l_k, l_{k'})$, the 1st, 7th, 13rd, 19th and 25th elements are $\text{Var}(l_k) \equiv$

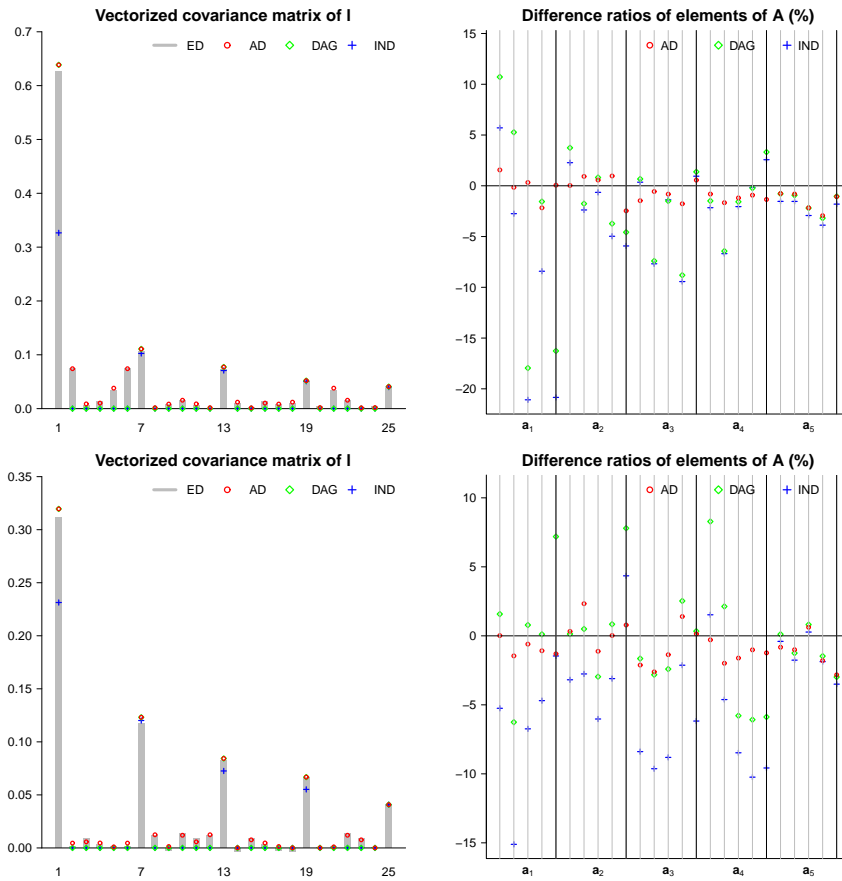


Fig. 1. The vectorized $\text{Cov}(l_k, l_{k'})$ and $\Delta^*(a_{kk'})$ for $k, k' = 1, \dots, 5$ for DGP 1 (Top pane) and DGP 2 (Bottom pane).

Table 2. The difference $\Delta(r_k)$ in percentage for $k = 1, \dots, 4$ for DGP 1 and DGP 2.

k	DGP1				DGP2			
	1	2	3	4	1	2	3	4
AD	-0.015	-0.011	-0.008	-0.000	0.009	0.010	-0.002	0.008
DAG	0.045	-0.009	0.031	0.027	0.009	0.030	0.013	0.020
IND	-0.125	-0.108	-0.018	0.010	-0.063	-0.016	-0.018	0.013

$\text{Cov}(l_k, l_k)$, and other elements are $\text{Cov}(l_k, l_{k'})$ for $k \neq k'$. From the figure, we can see AD fits ED well. IND tends to give a smaller $\text{Var}(l_k)$, while DAG gives the same $\text{Var}(l_k)$ as AD, both cases being consistent with the theoretical results. However, for $\text{Cov}(l_k, l_{k'}), k \neq k'$, IND and DAG claim that l_k are uncorrelated, which is not supported by ED.

For $\sigma(a_{kk'})$, AD performs pretty well and the difference ratios are controlled within 5%. In contrast, IND and DAG have poor performance; the difference ratio can even reach a figure as big as 20% for DGP 1.

As for the proportion of variation, Table 2 shows $\Delta(r_k)$ subscripted by AD, DAG and IND. Among the three asymptotic standard deviations, $\sigma_{\text{AD}}(r_k)$ is most close to $\sigma_{\text{ED}}(r_k)$, with the differences controlled within 0.02%. Although the differences for $\sigma_{\text{DAG}}(r_k)$ and $\sigma_{\text{IND}}(r_k)$ are larger, they are all controlled within 0.2%, which is negligible.

We have implemented two additional simulation DGPs, which is of higher order. See the Appendix A.2.

- **DGP 3:** VMA(2) model with multivariate Gaussian noise.
- **DGP 4:** VMA(3) model with multivariate Gaussian noise.

To summarize, for Gaussian processes, the simulation has confirmed the efficacy of Theorem 1 and indicated the deficiency of the diagonal spectral density matrix (DAG) assumption and the independent-data (IND) assumption for time series. On the positive side, the standard deviations of the proportion of variation have negligible difference under the three assumptions. Thus, making interval inference under which assumption makes not much difference for determining the number of effective PCs. However, DAG and IND

perform poorly for the eigenvectors, which can seriously affect the interpretation of PC loadings.

5.2. Efficacy of the estimation methods

In this part, we pay attention to the estimation of the standard deviation of the eigenvectors and proportions of variation, namely $\sigma(a_{kk'})$ and $\sigma(r_k)$. For each simulated sample of time length $n = 2000$ and dimension $p = 5$, we estimate the standard deviation by direct estimation. We average over $N = 2000$ samples and denote the average standard deviation as σ_{DE} . For each sample, we implement bootstrap with the MBB method. By reference to [Lahiri \(2003\)](#), which suggested that the optimal block size is of order $n^{1/3}$, we take the block size as 10, and the block number is 200 accordingly. Then, the bootstrap estimation of standard deviation of the eigenvectors are obtained with 500 bootstrap replications. We take average over $N = 2000$ samples and denote the average standard deviation as σ_{BE} .

Here we handle not only Gaussian processes, but also some non-Gaussian processes, including one process with outliers, one skewed process and two heavy-tailed processes. The models are as follows.

- **DGP 5:** VMA(1) model with multivariate Gaussian noise, but 1% of all the noises are outliers from other multivariate Gaussian distributions.
- **DGP 6:** VMA(1) model with noise from a multivariate skew normal distribution.
- **DGP 7:** VMA(1) model with noise from a multivariate $t(5)$ distribution.
- **DGP 8:** VMA(1) model with noise from a multivariate $t(8)$ distribution.

The coefficient matrix of the VMA(1) model is the same as that of DGP 2, check it and other details about the DGPs in [Appendix A.2](#).

For the proportion of variation, similar to [\(12\)](#), we can compute $\Delta_{\text{DE}}(r_k)$ and $\Delta_{\text{BE}}(r_k)$ with respect to ED for $k = 1, \dots, p$, which are shown in [Table 3](#) for DGPs 1-8. It can be seen that the bootstrap estimation behaves well except that it may be slightly sensitive to the outliers (DGP 5). As for the direct estimation, its behavior for heavy-tailed processes

Table 3. The $\Delta_{\text{DE}}(r_k)$ and $\Delta_{\text{BE}}(r_k)$ (%).

k	DGP1		DGP2		DGP3		DGP4	
	DE	BE	DE	BE	DE	BE	DE	BE
1	-0.013	-0.047	0.030	0.011	0.018	0.016	0.004	-0.024
2	0.021	-0.014	0.013	-0.000	0.019	0.010	0.021	-0.007
3	0.006	-0.010	0.014	-0.009	-0.002	-0.003	0.005	-0.006
4	0.017	0.003	0.012	-0.000	-0.001	-0.000	0.007	-0.004
k	DGP5		DGP6		DGP7		DGP8	
	DE	BE	DE	BE	DE	BE	DE	BE
1	0.035	0.252	-0.077	-0.015	-0.428	-0.088	-0.147	-0.041
2	0.016	0.182	-0.049	-0.012	-0.363	-0.067	-0.118	-0.022
3	0.026	0.137	-0.023	-0.010	-0.249	-0.028	-0.078	-0.010
4	0.006	0.044	-0.006	-0.006	-0.175	-0.038	-0.071	-0.020

(DGP 7 and 8) are worse than the bootstrap estimation. Nevertheless, all the values in the table are controlled within 0.5%, which is quite small in practice.

As for the loadings, similar to (13), we can compute $\Delta_{\text{DE}}^*(a_{kk'})$ and $\Delta_{\text{BE}}^*(a_{kk'})$ with respect to ED for $k, k' = 1, \dots, p$. Furthermore, to condense and summarize the element-wise information, we compute the mean absolute difference ratio ($\tilde{\Delta}$) for each eigenvector as

$$\tilde{\Delta}_{\text{DE}}(\mathbf{a}_k) = \sum_{k'=1}^p |\Delta_{\text{DE}}^*(a_{kk'})|/p, \quad \tilde{\Delta}_{\text{BE}}(\mathbf{a}_k) = \sum_{k'=1}^p |\Delta_{\text{BE}}^*(a_{kk'})|/p,$$

which can represent the accuracy of the standard deviation of the k th eigenvector. The $\tilde{\Delta}$ s (%) of PC loadings for DGPs 1-8 are shown in Table 4. From the table, it can be seen that for most models, the $\tilde{\Delta}$ s are controlled within 15%. Although some $\tilde{\Delta}$ s exceed 15%, such as in DGP 2 and 5, they occur in the less consequential 4th or 5th PC. For DGPs 1 to 6, most $\tilde{\Delta}$ s of DE are smaller than those of BE, meaning that the direct estimation method outperforms the bootstrap method. These models are Gaussian, skew normal or Gaussian with outliers, and are or are close to a Gaussian process. This fact might explain the success of the direct estimation method.

However, for the heavy-tailed processes in DGP 7 and 8, the direct estimation method

Table 4. The $\tilde{\Delta}_{DE}(\mathbf{a}_k)$ and $\tilde{\Delta}_{BE}(\mathbf{a}_k)$ (%).

	DGP1		DGP2		DGP3		DGP4	
	DE	BE	DE	BE	DE	BE	DE	BE
PC1	1.72	1.83	1.83	1.78	0.77	2.49	2.17	4.63
PC2	1.40	4.83	1.04	1.93	1.35	1.78	3.77	10.06
PC3	2.44	7.52	2.10	8.92	1.01	1.73	0.82	4.72
PC4	7.32	10.38	9.36	17.58	1.01	1.79	2.00	4.47
PC5	4.87	5.10	7.77	11.07	1.13	1.15	1.32	0.88
	DGP5		DGP6		DGP7		DGP8	
	DE	BE	DE	BE	DE	BE	DE	BE
PC1	1.89	9.11	1.34	1.44	29.02	3.19	11.70	2.07
PC2	1.86	8.42	1.51	1.33	29.66	1.36	11.11	1.81
PC3	3.19	12.22	1.25	2.33	32.39	6.07	14.65	4.74
PC4	17.45	13.09	2.34	7.82	30.02	12.08	13.47	9.84
PC5	1.12	7.12	1.45	9.46	36.47	11.00	19.48	5.08

behaves poorly. The tail of DGP 7 is heavier than DGP 8, and the performance is worse. The poor performance for the heavy-tailed process is due to the joint fourth cumulant of heavy-tailed processes being non-negligible. Thus, we recommend the bootstrap method, which can be implemented for general processes easily.

6. An empirical example

Now, we return to principal portfolio management by studying an empirical example, the PCA of which plays a central role, so as to illustrate the importance of Theorem 1 and the utility of the bootstrap method, and to sound a warning against cavalier usage of PCA in time series without due attention to the standard errors of the estimated eigenvectors. In doing so, our hope is that our procedure will provide a paradigm of implementing PCA for time series that encompasses estimation, inference and interpretation.

We use a set of the daily returns of 10 stocks traded in the New York Stock Exchange. It consists of CVX (Chevron), XOM (Exxon), AAPL (Apple), FB (Facebook), MSFT

(Microsoft), MRK (Merck), PFE (Pfizer), BAC (Bank of America), JPM (JP Morgan), and WFC (Wells Fargo & Co.). The data are from August 2, 2016 to December 30, 2016, a total of 106 days. Within the 10 stocks, CVX and XOM belong to the energy industry, while AAPL, FB and MSFT are in the information technology (IT) sector. MRK and PFE are both health companies. BAC, JPM, and WFC are in the financial sector. See, e.g., Section 4.5 of [Wei \(2019\)](#) for a fuller description of the data set.

First, we describe the data and check informally whether the data satisfy the basic assumptions: ergodicity, stationarity and Gaussianity. From the marginal distribution, the time series plot in [Fig.2](#) and the QQ plot of the normalized data in [Fig.3](#), we can see that the time series do not appear to display obvious trend non-stationarity apart from a few possible outliers. The data appear to be non-Gaussian, and there are heavy tails. Thus, we consider using the bootstrap method to estimate the standard deviations, based on which we can make inference on the number and loadings of the PCs.

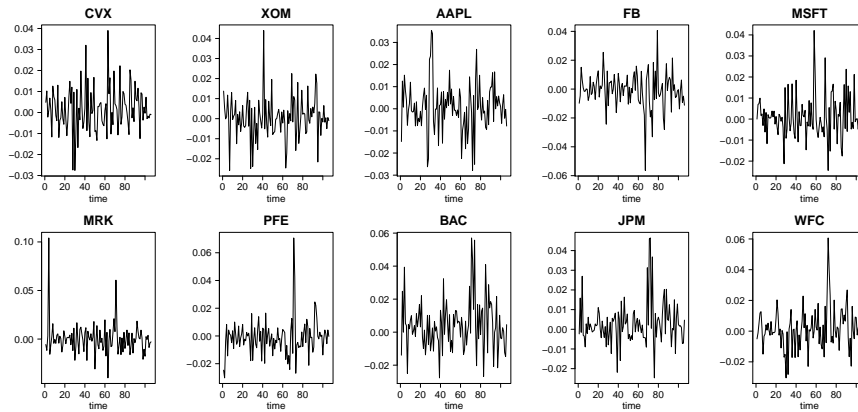


Fig. 2. The time series plot of the stock returns data.

In [Fig.4](#), we plot the proportion of variation and its 95% confidence interval with the estimated standard deviation by the bootstrap method. The eigenvalues of the first four components account for $r_4 = 77.8\%$ of the total sample variance. This value together with a visual inspection of [Fig.4](#) suggests the choice of four PCs. (In [Wei \(2019\)](#), the first four components account for 77.5%, and the small difference is due to a recording error of the fourth eigenvalue in [Wei \(2019\)](#). He recorded 0.00011 but re-running his R-code we

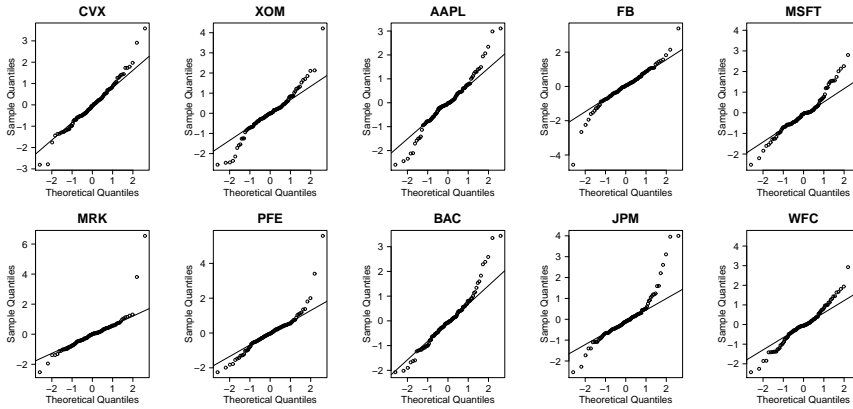


Fig. 3. The QQ plot of the normalized stock returns data.

obtained 0.00013. Our value, but not his, is consistent with Figure 4.2 in the book.)

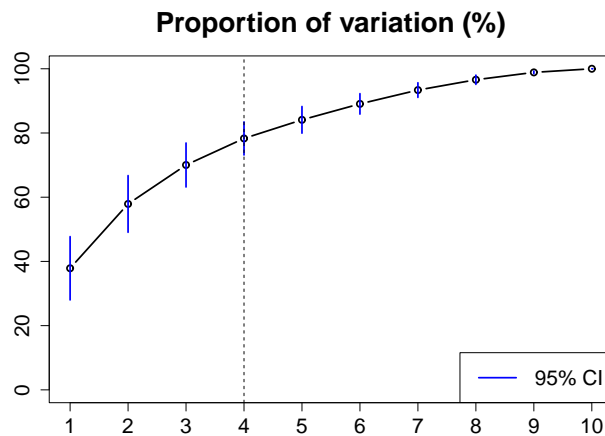


Fig. 4. The proportion of variation with 95% confidence interval.

The PC loadings are shown in Table 5. The dash-lines in the tables separate the sectors. For every element of the first four PC loadings, $a_{kk'}$ for $k = 1, \dots, 4, k' = 1, \dots, 10$, we test the null hypothesis $H_0 : a_{kk'} = 0$ with the estimated standard deviations in Theorem 1 by the bootstrap method. To compare, we conduct similar tests under the DAG assumption (Corollary 3) and under the IND assumption (Corollary 4), using the estimated standard deviations by the direct estimation method. The PC loadings and inference are shown

Table 5. The PC loadings.

	CVX	XOM	AAPL	FB	MSFT	MRK	PFE	BAC	JPM	WFC
PC1	0.166	0.230	0.028	0.088	0.075	0.470	0.369	0.539	0.391	0.324
PC2	0.104	0.126	0.431	0.513	0.362	0.329	0.167	-0.289	-0.175	-0.379
PC3	0.407	0.243	0.173	0.319	0.286	-0.596	-0.316	0.191	0.088	0.251
PC4	0.576	0.552	-0.354	-0.231	-0.197	0.035	0.126	-0.222	-0.091	-0.265
PC5	0.120	0.026	0.180	-0.324	0.127	0.396	-0.725	0.279	0.012	-0.266
PC6	0.034	0.203	0.377	-0.213	-0.116	0.223	-0.118	-0.438	-0.206	0.678
PC7	0.016	-0.088	-0.639	0.505	0.052	0.320	-0.358	-0.164	-0.065	0.255
PC8	0.099	-0.023	0.264	0.399	-0.841	0.024	-0.136	0.155	-0.009	-0.097
PC9	0.622	-0.686	0.069	-0.093	0.011	0.065	0.080	-0.217	0.265	0.041
PC10	0.227	-0.215	-0.042	-0.033	0.051	0.031	0.165	0.416	-0.827	0.110

in Table 6, where insignificant elements are suppressed (at 5% significant level), and significant elements are shown with their signs. The values in Table 5 are the same as that shown in Wei (2019) (after reversing the signs of some loadings to be consistent with our results). He did not conduct inference to the loadings but did a rough screening by an arbitrary truncation in that loadings smaller than 0.1 (in absolute values) are suppressed, which are also shown in Table 6.

As we can see from the significant loadings in Table 6, our Theorem 1 leads to the fewest number of significant elements while the arbitrary truncation in Wei (2019) gives the largest number. The results of DAG and IND are almost identical, with DAG making minimal impact. The IND is of particular interest because it represents to-date the most frequently (mis-)used PCA for time series data, i.e., ignoring the dependence and non-Gaussianity of the observed data. Now, the significant loadings in Table 6 lead to different interpretations of the first four PCs, as summarized in Table 7. The result based on Theorem 1 tends to give a sharper identification of important stocks. Note also that with Theorem 1 significant stocks always lie within the same sector, while this is not the case with IND (Corollary 4) or with Wei (2019). The above observations are particularly relevant for principal portfolios management mentioned in Section 1 because they can help the investors to formulate more effectively their investment strategies on which stocks to long or short. Specifically, it is interesting to point out that PC2 and PC3 based on Theorem 1 are particularly relevant due to the clear contrast between the IT sector and the finance sector in PC2, and that between the IT sector and the health sector in PC3.

Table 6. The PC loadings with inference from Theorem 1, DAG (Corollary 3), IND(Corollary 4), and truncation in Wei (2019).

		CVX	XOM	AAPL	FB	MSFT	MRK	PFE	BAC	JPM	WFC
Theorem 1	PC1	+	+				+	+	+	+	+
	PC2			+	+	+				-	-
	PC3					+	-	-			
	PC4	+	+								
DAG (Corollary 3)	PC1	+	+				+	+	+	+	+
	PC2			+	+	+	+		-	-	-
	PC3	+			+	+	-	-	+		+
	PC4	+	+	-							
IND (Corollary 4)	PC1	+	+				+	+	+	+	+
	PC2			+	+	+	+	+	-	-	-
	PC3	+			+	+	-	-	+		+
	PC4	+	+	-					-		
Truncation (Wei, 2019)	PC1	+	+				+	+	+	+	+
	PC2	+	+	+	+	+	+	+	-	-	-
	PC3	+	+	+	+	+	-	-	+		+
	PC4	+	+	-	-	-		+	-		-

Table 7. Interpretations of the first four PCs based on Theorem 1 and IND (Corollary 4).

	Theorem 1	IND (Corollary 4)	Wei (2019)
PC1	The general market other than IT.	The general market other than IT.	The general market other than IT.
PC2	Financial vs. IT.	Financial vs. IT & health.	Financial vs. non-financial.
PC3	Health vs. IT.	Health vs. non-health.	Health vs. non-health.
PC4	Energy.	Energy vs. IT & financial.	Energy vs. non-energy.

Accordingly, the IT sector seems to be set well apart from other sectors, indicating different positions (long or short) from other sectors in the portfolio. The different interpretations demonstrate the importance of an appropriate significance test of the loadings.

We argue that without an appropriate test, users of PCA may be unaware of the subtle variabilities of loadings, and tend to mis-interpret the eigenvectors.

7. Conclusion

In this paper, by building on [Hannan \(1976\)](#), we have made explicit the asymptotic sampling properties of the eigenvalues and eigenvectors in PCA for stationary and ergodic multivariate linear time series and assessed their efficacy. The classical PCA results for independent data and the Taniguchi-Krishnaiah method are shown to be deficient for time series. Although they have led to simpler theoretical results, simulations have revealed their poor performance for time series data, especially for the eigenvectors. Thus, while the number of PCs is quite robust in respect of different dependence structure assumption, the interpretation of the principal component loadings may be questionable under these assumptions. Direct and bootstrap method to estimate the asymptotic covariance have been given to draw inference in practice. We argue that when applying a PCA to time series, the dependence of data should not be ignored and we have shown step by step how an appropriate inference can be conducted.

Essentially, the shortcomings of the two defective methods for time series lie with the fact that they have ignored some information contained in cross covariances. Another way to exploit the omitted information is to adopt a frequency-domain PCA. See, e.g., [Brillinger \(2001\)](#) and [Priestley, Subba Rao and Tong \(1974\)](#). However, unsolved problems remain with this approach, the possible non-causality of the principal components being a major one. They await further research.

Acknowledgements

We are most grateful to the Joint Editor and the referees for their constructive comments and suggestions.

A. Appendix

In this appendix, we give more information of the simulation, including the details of DGPs and the results of the two additional simulation in Section 5.1.

A.1. Details of the DGPs

The coefficient matrices of the DGPs are shown as follows.

DGP 1: VAR(1).

$$\mathbf{F}(1) = \begin{pmatrix} 0.21 & -0.49 & 0.16 & -0.14 & -0.36 \\ -0.25 & 0.074 & 0.38 & -0.14 & 0.047 \\ -0.11 & 0.26 & 0.39 & 0.091 & 0.18 \\ -0.41 & 0.37 & 0.066 & 0.37 & 0.028 \\ 0.46 & -0.46 & 0.094 & 0.18 & -0.41 \end{pmatrix}.$$

DGP 2, 5-8: VMA(1).

$$\mathbf{G}(1) = \begin{pmatrix} -0.46 & 0.17 & 0.23 & 0.40 & -0.22 \\ 0.15 & -0.59 & 0.05 & -0.26 & -0.24 \\ 0.13 & -0.32 & -0.26 & -0.28 & -0.41 \\ 0.15 & 0.20 & 0.51 & -0.38 & -0.55 \\ 0.43 & 0.02 & -0.25 & -0.32 & -0.34 \end{pmatrix}.$$

DGP 3: VMA(2).

$$\mathbf{G}(1) = \begin{pmatrix} 0.62 & -0.73 & -0.02 & 0.52 & -0.52 \\ 0.71 & -0.11 & -0.19 & 0.43 & -0.75 \\ 0.09 & -0.56 & 0.79 & -0.51 & -0.08 \\ 0.14 & 0.90 & 0.07 & -0.53 & -0.27 \\ -0.23 & -0.01 & 0.59 & 0.81 & 0.10 \end{pmatrix},$$

$$\mathbf{G}(2) = \begin{pmatrix} 0.26 & -0.16 & -0.16 & -0.38 & 0.27 \\ 0.34 & -0.27 & -0.33 & -0.44 & 0.06 \\ -0.22 & -0.18 & 0.62 & -0.08 & 0.19 \\ 0.26 & -0.30 & 0.15 & -0.24 & 0.02 \\ 0.11 & 0.21 & 0.21 & -0.14 & 0.45 \end{pmatrix}.$$

DGP 4: VMA(3).

$$\mathbf{G}(1) = \begin{pmatrix} 0.94 & -0.35 & -0.49 & 0.17 & -0.18 \\ 0.58 & 0.35 & -0.43 & -0.29 & -0.36 \\ 0.42 & -0.16 & 1.07 & -0.28 & 0.39 \\ 0.59 & 0.41 & -0.38 & 0.27 & -0.03 \\ 0.18 & 0.66 & -0.28 & 0.42 & 0.91 \end{pmatrix},$$

$$\mathbf{G}(2) = \begin{pmatrix} -0.14 & -0.04 & -0.33 & 0.40 & 0.02 \\ 0.27 & -0.32 & -0.27 & 0.02 & -0.16 \\ 0.45 & -0.18 & 0.03 & -0.31 & 0.16 \\ 0.41 & 0.39 & -0.33 & -0.36 & 0.10 \\ 0.31 & 0.47 & -0.08 & 0.14 & -0.26 \end{pmatrix},$$

$$\mathbf{G}(3) = \begin{pmatrix} -0.16 & 0.24 & 0.10 & 0.22 & 0.17 \\ -0.22 & 0.17 & 0.10 & 0.26 & 0.14 \\ 0.06 & -0.03 & -0.11 & -0.05 & -0.17 \\ -0.11 & 0.02 & 0.01 & 0.19 & 0.11 \\ 0.14 & -0.11 & 0.16 & -0.21 & -0.24 \end{pmatrix}.$$

As for the distributions of the noise, those of DGP 1-4 are shown in Section 5. Now we give details and parameters of the noise distribution of DGP 5-8.

DGP 5: VMA(1) model with multivariate Gaussian noise, but 1% of all the noises are outliers from another multivariate Gaussian distribution.

First we draw 2000 observations of $\mathbf{e}(t)$ from $\mathcal{N}(0_5, 10\mathbf{I}_5)$. Then, 20 random numbers from 1 to 2000 are drawn, which will be the positions of the outliers. The original observations in the first 10 positions are replaced by random vectors from $\mathcal{N}(10_5, 10\mathbf{I}_5)$, and those in the last 10 positions are replaced by random vectors from $\mathcal{N}(-10_5, 10\mathbf{I}_5)$.

DGP 6: VMA(1) model with noise from a multivariate skew normal distribution.

The multivariate skew normal distribution is discussed by [Chan and Tong \(1986\)](#) and [Azzalini and Capitanio \(1999\)](#) and here we use the parameterization in the latter, $\mathcal{SN}_5(\xi, \Omega, \alpha)$. We set $\xi = 0_5$, $\Omega = 10\mathbf{I}_5$ and $\alpha = (1, 2, 3, 4, 5)$.

DGP 7: VMA(1) model with noise from a multivariate $t(5)$ distribution.

DGP 8: VMA(1) model with noise from a multivariate $t(8)$ distribution.

For the multivariate t -distribution $t_v(\mu, \Sigma)$, we set $\mu = 0_5$, $\Sigma = 10I_5$, $v = 5$ for DGP 7 and $v = 8$ for DGP 8.

A.2. Additional simulation

In this subsection, we present two additional simulations, namely DGP 3 and 4, to enrich Subsection 5.1 and assess the efficacy of the asymptotics for Gaussian processes.

For each model, the vectorized $\text{Cov}(l_k, l_{k'})$ and $\Delta^*(a_{kk'})$ in (13) for $k, k' = 1, \dots, 5$ are each a 25-dimensional vector. See their plots in Fig.5 for DGP 3 and 4 respectively. For

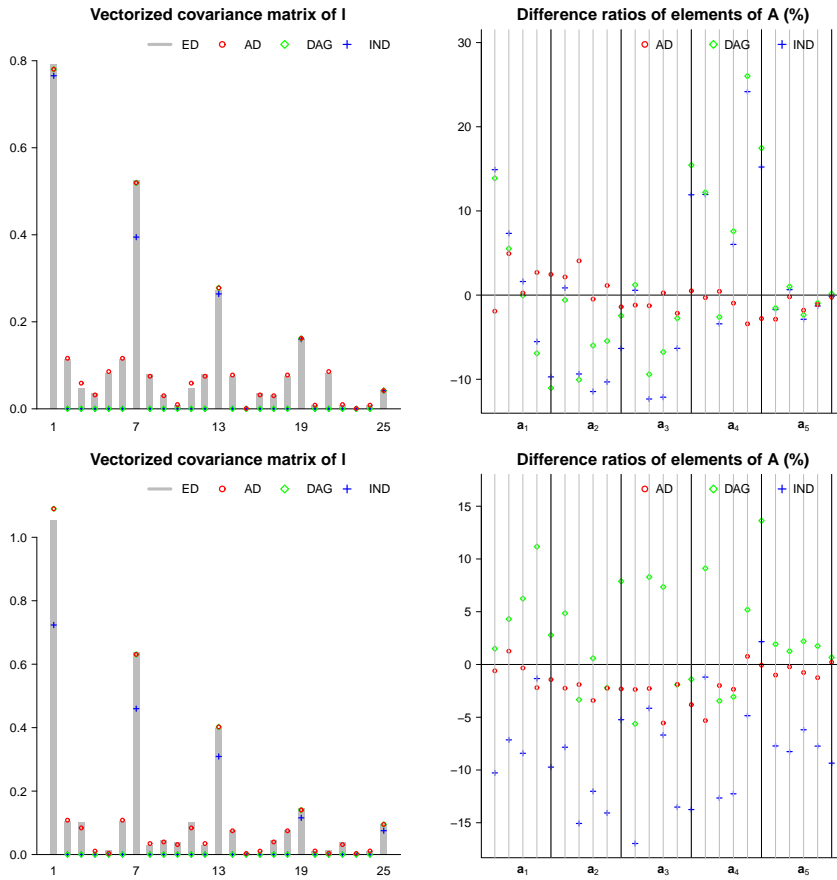


Fig. 5. The vectorized $\text{Cov}(l_k, l_{k'})$ and $\Delta(a_{kk'})$ for $k, k' = 1, \dots, 5$ for DGP 3 (Top pane) and DGP 4 (Bottom pane).

the vectorized $\text{Cov}(l_k, l_{k'})$, we can see AD fits ED well. IND tends to give a smaller

Table 8. The difference $\Delta(r_k)$ in percentage for $k = 1, \dots, 4$ for DGP 3 and DGP 4.

k	DGP1				DGP2			
	1	2	3	4	1	2	3	4
AD	-0.007	-0.000	0.000	0.000	0.011	0.010	0.003	0.003
DAG	0.044	0.026	0.043	0.020	0.035	0.011	0.018	0.005
IND	0.031	0.008	0.035	0.018	-0.068	-0.069	-0.028	-0.023

$\text{Var}(l_k)$, while DAG gives the same $\text{Var}(l_k)$ as AD, both cases being consistent with the theoretical results. However, for $\text{Cov}(l_k, l_{k'}), k \neq k'$, IND and DAG claim that l_k are uncorrelated, which is not supported by ED.

For $\sigma(a_{kk'})$, AD performs pretty well and the difference ratios are controlled within 6%. In contrast, IND and DAG have poor performance; the difference ratio can even reach a figure as big as 20% for DGP 3.

As for the proportion of variation, Table 8 shows $\Delta(r_k)$ subscripted by AD, DAG and IND. Among the three asymptotic standard deviations, $\sigma_{\text{AD}}(r_k)$ is most close to $\sigma_{\text{ED}}(r_k)$, with the differences controlled within 0.02%. Although the differences for $\sigma_{\text{DAG}}(r_k)$ and $\sigma_{\text{IND}}(r_k)$ are larger, they are all controlled within 0.1%, which is negligible.

Supplementary material

Supplementary material includes R codes for the empirical example in Section 6.

References

- AHAMAD, B. (1967). An analysis of crimes by the method of principal components. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **16**(1) 17–35.
- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ.
- AZZALINI, A. and CAPITANIO, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61**(3) 579–602.

- BRILLINGER, D. R. (2001). *Time Series: Data Analysis and Theory*. SIAM, Philadelphia, PA.
- BÜHLMANN, P. (2002). Bootstraps for time series. *Statist. Sci.* **17**(1) 52–72.
- CARLSTEIN, E. (1986). Simultaneous confidence regions for predictions. *Amer. Statist.* **40**(4) 277–279.
- CHAN, K. S. and TONG, H. (1986). A note on certain integral equations associated with non-linear time series analysis. *Probab. Theory Relat. Fields.* **73**(1) 153–158.
- CRADDOCK, J. M. (1965). A meteorological application of principal component analysis. *Journal of the Royal Statistical Society. Series D (The Statistician)*. **15**(2) 143–156.
- FEENEY, G. J. and HESTER, D. D. (1967). Stock market indices: A principal components analysis. In *Risk Aversion and Portfolio Choice* (D. D. Hester and J. Tobin, eds.) 110–138. Wiley, New York.
- FLURY, B. (1997). *A First Course in Multivariate Statistics*. Springer, New York.
- GIRAITIS, L. and ROBINSON, P. M. (2001). Whittle estimation of ARCH models. *Econometric Theory*. **17**(3) 608–631.
- HALL, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.* **20**(2) 231–246.
- HANNAN, E. J. (1976). The asymptotic distribution of serial covariances. *Ann. Statist.* **4**(2) 396–399.
- JOLLIFFE, I. T. (2002). *Principal Component Analysis*, 2nd ed. Springer, New York.
- KEIJRSERS, B. and VAN DIJK, D. (2019). Uncertainty and the macro-economy: An out-of-sample investigation. Unpublished report. Tinbergen Institute, the Netherlands.
- KIRCH, C. and POLITIS, D. N. (2011). TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *Ann. Statist.* **39**(3) 1427–1470.
- KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**(3) 1217–1241.

- LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York.
- LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (LePage, R. and Billard, L., eds.) 225–248. Wiley, New York.
- MARYON, R. H. (1979). Eigenanalysis of the Northern Hemispherical 15-day mean surface pressure field and its application to long-range forecasting. *Met. O 13 Branch Memorandum No. 82* (Unpublished). U.K. Meteorological Office, Bracknell.
- MEYER, M., PAPANODITIS, E. and KREISS, J.-P. (2020). Extending the validity of frequency domain bootstrap methods to general stationary processes. *Ann. Statist.* **48**(4) 2404–2427.
- PARTOVI, M. H. and CAPUTO, M. (2004). Principal portfolios: recasting the efficient frontier. *Econ. Bull.* **7**(3) 1–10.
- PASINI, G. (2017). Principal component analysis for stock portfolio management. *Int J Pure Appl Math.* **115**(1) 153–167.
- POLITIS, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statist. Sci.* **18**(2) 219–230.
- PRIESTLEY, M., SUBBA RAO, T. and TONG, H. (1974). Applications of principal component analysis and factor analysis in the identification of multivariable systems. *IEEE Trans. Automat. Control.* **19**(6) 730–734.
- SHAO, X. (2009). Confidence intervals for spectral mean and ratio statistics. *Biometrika.* **96**(1) 107–117.
- STONE, R. (1947). On the interdependence of blocks of transactions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **9**(1) 1–45.
- TANIGUCHI, M. and KRISHNAIAH, P. (1987). Asymptotic distributions of functions of the eigenvalues of sample covariance matrix and canonical correlation matrix in multivariate time series. *J. Multivariate Anal.* **22**(1) 156–176.

- TANIGUCHI, M., PURI, M. L. and KONDO, M. (1996). Nonparametric approach for non-Gaussian vector stationary processes. *J. Multivariate Anal.* **56** 259–283.
- THEILER, J., EUBANK, S., LONGTIN, A., GALDRIKIAN, B. and FARMER, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Phys. D.* **58**(1–4) 77–94.
- TSAY, R. S. (2005). *Analysis of Financial Time Series*. 2nd ed. Wiley, Hoboken, NJ.
- TSAY, R. S. (2014). *Multivariate Time Series Analysis: With R and Financial Applications*. Wiley, Hoboken, NJ.
- WEI, W. W. (2019). *Multivariate Time Series Analysis and Applications*. Wiley, Hoboken, NJ.
- YANG, L. (2015). An application of principal component analysis to stock portfolio management. *Department of Economics and Finance, University of Canterbury*.