I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2014/31

Large-Sample Approximations for Variance-Covariance Matrices of High-Dimensional Time Series

STELAND, A. and R. VON SACHS

LARGE-SAMPLE APPROXIMATIONS FOR VARIANCE-COVARIANCE MATRICES OF HIGH-DIMENSIONAL TIME SERIES

Ansgar Steland and Rainer von Sachs

Institute of Statistics
RWTH Aachen University
Wüllnerstr. 3, D-52056 Aachen, Germany
steland@stochastik.rwth-aachen.de
and
Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA)
Université catholique de Louvain
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
rainer.vonsachs@uclouvain.be

July 15, 2014

Distributional approximations of (bi–) linear functions of sample variance-covariances matrices play a critical role to analyze vector time series, as they are needed for various purposes, especially to draw inference on the dependence structure in terms of second moments and to analyze projections onto lower dimensional spaces as those generated by principal components. This particularly applies to the high-dimensional case, where the dimension d is allowed to grow with the sample size n and may even be larger than n. We establish large-sample approximations for such bilinear forms related to the sample variance-covariance matrix of a high-dimensional vector time series in terms of strong approximations by Brownian motions. The results cover weakly dependent as well as many long-range dependent linear processes and are valid for uniformly ℓ_1-bounded projection vectors, which arise, either naturally or by construction, in many statistical problems extensively studied for high-dimensional series. Among those problems are sparse financial portfolio selection, sparse principal components, shrinkage estimation and change-point analysis for high–dimensional time series, which are discussed in greater detail.

AMS 2000 subject classifications: Primary 60F17, 62E20, 62H10; secondary 60F05, 62L10, 62H25.

Keywords: Brownian motion, change-point, empirical process, principal components, shrinkage, sparsity, strong approximation.

1. INTRODUCTION

The estimation of high-dimensional variance-covariance matrices based on a vector time series arises in diverse areas such as financial portfolio optimization, image analysis and multivariate time series analysis in general. Of particular interest is the case that the dimension $d = d_n$ of the time series grows even faster than the sample size n. Due to the lack of consistency of the sample variance-covariance estimator with respect to commonly used norms such as the Frobenius norm, various regularized modifications have been proposed and extensively studied within a high-dimensional context. For example, banding and tapering estimators, recently studied by [2] for Gaussian samples, may achieve consistency if $\log d_n/n = o(1)$. The performance of shrinkage estimation, a widely used technique dating back to the seminal work of [26], has been investigated by [19] for i.i.d. samples of growing dimension and further studied for the weak dependent case in [23]. For results on shrinkage estimation in the frequency domain we refer to [4].

However, often the estimation of the d_n^2-dimensional variance-covariance matrix is an intermediate step and one is mainly interested in the behavior of functions of the sample variance-covariance matrix, especially quadratic and bilinear forms which naturally arise when studying projection type statistics. In addition, one often needs distributional approximations of such functions, in order to construct statistical decision procedures. Whereas consistency and performance properties have been already investigated to some extent, there are only a few results about the asymptotic distribution theory in the sense of distributional convergence (weak convergence) and strong approximations by Brownian motions, respectively, going beyond the classical results. To the best of our knowledge, there are no results addressing sample estimates of (auto–) covariances within a high-dimensional framework for correlated time series. In some sense close to the present paper are the following results for fixed dimension. [36] derived a CLT for a finite number of sample auto-covariances assuming a linear process. [18], also working within a linear process framework, established large-sample distributional asymptotics, based on strong approximations, of the sample cross-covariance matrix for two time processes. His assumptions are weak enough to cover not only the weak dependent case, but also a wide range of long-range dependent processes.

The present paper builds upon the latter result by establishing strong approximations of bilinear forms associated to the centered sample covariance matrix of a high-dimensional vector time series. The result implies the validity of a central limit theorem (CLT) for scaled bilinear forms $\sqrt{n}\mathbf{v}_n'[\widehat{\boldsymbol{\Sigma}}_n - E(\widehat{\boldsymbol{\Sigma}})]\mathbf{w}_n$, where $\widehat{\boldsymbol{\Sigma}}_n$ is the usual sample variance-covariance matrix and $\mathbf{v}_n$ and $\mathbf{w}_n$ are weighting vectors. It turns out that d_n may even grow faster than n.

Concerning the weighting (or projection) vectors, our results assume that they are uniformly bounded in the ℓ_1-norm. Such projections naturally arise in many problems studied in the area of high-dimensional statistics and probability: Sparse optimal portfolio selection, as recently studied by [5], deals with explicit construction of ℓ_1-bounded portfolios from historical data sets. The same applies to several approaches of sparse principal component analysis, especially those of [15], [24], [35] and [34], where ℓ_1-bounded principal components are constructed, in order represent high-dimensional data by only a few projections. We discuss those applications in greater detail in Section 5. We also illustrate how the results can be applied to obtain distributional approximations of shrinkage estimators of a high-dimensional covariance matrix. Lastly, we discuss the application to change detection procedures. Such procedures aim at detecting a change in the distribution and have been thoroughly studied for various second order problems for time series observations. Of course, a change in a covariance $\gamma_X(i,j) = E(X^{(i)}X^{(j)})$, $t \geq 1$, of a vector time series can be detected by applying a method to the sequence $X_t^{(i)}X_t^{(j)}$, $t \geq 1$, which is sensitive to location changes such as those discussed in [12], [1], detectors based on local linear estimators proposed by [29], kernel detectors as studied by [27] and [28], or methods based on characteristic functions, see e.g. [31] and the references therein.

The organization of the paper is as follows. Section 2 explains the general setting, discusses its basic relationship to projection-based analyses and introduces the bilinear form of interest. Notation, the specific model for the vector time series and its interpretation in terms of an infinite–dimensional latent factor model as well as assumptions are introduced and discussed in Section 3. Section 4 provides the strong approximation with rate, its proof and related FCLTs and CLTs. Lastly, Section 5 elaborates on several statistical problems to which our results are directly applicable.

2. PROJECTION-BASED ANALYSIS OF HIGH-DIMENSIONAL TIME SERIES

Let us assume that we observe d possibly dependent time series such that at time n we are given the observations

$$Y_i^{(\nu)}, \dots, Y_i^{(\nu)}, \qquad \nu = 1, \dots, d, \ 1 \leq i \leq n,$$

where the dimension $d = d_n$ may grow with the sample size n, such that, as time proceeds, there may be more and more time series available. Equivalently, we are given a time series of length n of possibly dependent random vectors

$$\mathbf{Y}_{ni} = (Y_i^{(1)}, \dots, Y_i^{(d_n)})', \qquad 1 \leq i \leq n,$$

of dimension d_n, constituting the $(n \times d_n)$–dimensional data matrix

$$\mathcal{Y}_n = \left(Y_i^{(j)} \right)_{1 \leq i \leq n, 1 \leq j \leq d_n}.$$

We are interested in the second moment structure and thus assume $E(Y_n^{(j)}) = 0$ for all $i = 1, \dots, n$ and $n \geq 1$. Our assumptions on the coordinate processes, basically that they are linear processes with sufficiently fast decreasing coefficients, are weak enough to cover the common framework of correlated ARMA(p,q)-processes and also allow for a wide class of long-range dependent series.

Let us assume for a moment that $\mathbf{Y}_{n1}, \dots, \mathbf{Y}_{nn}$ is stationary, a condition that we shall relax later, and let

$$\mathbf{Y}_n = (Y^{(1)}, \dots, Y^{(d_n)})'$$

be a generic copy. For the analysis of such highdimensional time series, the unknown variance-covariance matrix

$$\mathbf{\Sigma}_n = E(\mathbf{Y}_n \mathbf{Y}_n') = \left(E(Y^{(\nu)}Y^{(\mu)}) \right)_{1 \leq \nu, \mu \leq d_n}$$

is of substantial interest, but difficult to estimate from past data, in particular if $d_n >> n$. It comprises the second-order information on the dependence structure of the d_n variables. Thus, any conclusions on the correlation structure has to rely on estimators calculated from the time series, and inferential procedures require appropriate large-sample asymptotics. Let

$$\widehat{\mathbf{\Sigma}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_{ni}\mathbf{Y}_{ni}' \tag{2.1}$$

be the $(d_n \times d_n)$-dimensional sample variance–covariance matrix with elements

$$\widehat{\sigma}_{\nu\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i^{(\nu)} Y_i^{(\mu)}, \qquad \nu, \mu = 1, \dots, d_n.$$

Before proceeding, let us observe that the above framework also covers the case of a univariate time series $\{Z_k : k \geq 0\}$ as an interesting special case. The *embedding* is given by

$$\mathbf{Y}_{ni} = (Z_i, Z_{i+1}, \dots, Z_{i+d_n-1})', \qquad i = 1, \dots, n.$$

Then

$$\widehat{\sigma}_{\nu\mu} = \frac{1}{n}\sum_{i=1}^{n} Z_{i+\nu-1} Z_{i+\mu-1}, \qquad 1 \leq \nu, \mu \leq d_n.$$

It follows that the $(h+1)$th element of the first row of $\widehat{\mathbf{\Sigma}}_n$ estimates $\gamma_Z(h) = E(Z_0 Z_h)$ using the observations $Z_1, \dots, Z_T$, where $T = n + h$, and can be written as

$$\widetilde{\gamma}_Z(h) = \frac{T}{T-h}\widehat{\gamma}_Z(h), \qquad \widehat{\gamma}_Z(h) = \frac{1}{T}\sum_{i=1}^{T-h} Z_i Z_{i+h},$$

for $h = 0, \dots, T-1$. In a similar way, one may consider autocovariances and cross-covariances of, say r_n time series $\{Z_k^{(l)} : k \geq 0\}$, $l = 1, \dots, r_n$.

Estimators of $\mathbf{\Sigma}_n$ are also needed and have to be evaluated in terms of their asymptotic laws, when interest focuses on the analysis of (a set of) linear combinations of $\mathbf{Y}_n$. Typical examples are convex combinations, contrasts and, more generally, projections. Thus let

$$\mathbf{w}_n = (w_1, \dots, w_{d_n})', \qquad n \geq 1,$$

be a sequence of weights $w_j = w_{d_n j}$, not necessarily non-negative, with uniformly bounded ℓ_1-norm, i.e.

$$\sup_{n\in\mathbb{N}} \|\mathbf{w}_n\|_{\ell_1} = \sup_{n\in\mathbb{N}} \sum_{\nu=1}^{d_n} |w_j| < \infty \tag{2.2}$$

The variance of the projection $\mathbf{w}_n'\mathbf{Y}_n$ is given by $\mathbf{w}_n'\mathbf{\Sigma}_n\mathbf{w}_n$ and can be estimated nonparametrically by the quadratic form

$$\mathbf{w}_n'\widehat{\mathbf{\Sigma}}_n\mathbf{w}_n,$$

whose random fluctuations may severely affect any inferential procedure related to $\mathbf{w}_n'\mathbf{Y}_n$. Recalling that, especially in a high-dimensional setting, functions depending on a large number of covariances tend to be difficult to estimate accurately, distributional approximations for weighted sums of subsets $\widehat{\mathbf{\Sigma}}_{\mathcal{I},\mathcal{J}} = \{\widehat{\sigma}_{ij} : i \in \mathcal{I}, j \in \mathcal{J}\}$ of $\widehat{\mathbf{\Sigma}}_n$, where $\mathcal{I}, \mathcal{J} \subset \{1, \dots, d_n\}$, are of interest. Thus, we study, more generally, the bilinear form

$$Q_n(\mathbf{v}_n, \mathbf{w}_n) = \mathbf{v}_n'\widehat{\mathbf{\Sigma}}_n\mathbf{w}_n$$

for weighting vectors $\mathbf{v}_n$ and $\mathbf{w}_n$ with uniformly bounded ℓ_1-norms. Observe that $Q_n(\mathbf{v}_n, \mathbf{w}_n)$ also provides us with an estimator of the covariance of the projections $\mathbf{v}_n'\mathbf{Y}_n$ and $\mathbf{w}_n'\mathbf{Y}_n$.

It is worth discussing the uniform ℓ_1-condition, which arises in various applications such as optimal porfolio selection and sparse principal component analyses as those proposed by [15], [34] and [35]. It is also worth mentioning that Q_n remains bounded even for degenerate covariance matrices, if $\mathbf{v}_n$ and $\mathbf{w}_n$ have uniformly bounded ℓ_1-norm, as can be seen if we put $\mathbf{\Sigma}_n = \sigma\mathbf{1}\mathbf{1}'$, where here and throughout the article $\mathbf{1} = (1, \dots, 1)' \in \mathbb{R}^{d_n}$, leading to

$$|Q_n(\mathbf{v}_n, \mathbf{w}_n)| = \sigma|\mathbf{v}_n'\mathbf{1}\mathbf{1}'\mathbf{w}_n| = \sigma|\sum_i v_{ni} \sum_i w_{ni}| \le \sigma\|\mathbf{v}_n\|_{\ell_1}\|\mathbf{w}_n\|_{\ell_1}.$$

Thus, the ℓ_1-norm condition is a natural one. Especially, it ensures that Q_n maps products $\mathcal{U}_\delta \times \mathcal{U}_\delta$ of δ-balls

$$\mathcal{U}_\delta = \{\{\mathbf{v}_n\} : \sup_{n\in\mathbb{N}} \|\mathbf{v}_n\|_{\ell_1} \le \delta\},$$

for $\delta > 0$, onto bounded sets, for all covariance matrices with uniformly bounded entries, thus including cases that correspond to perfectly correlated coordinates.

The behavior of projections for high-dimensional observations has also been studied by [9], but from a different perspective. There it is shown that for large dimension d projections $\mathbf{w}'\mathbf{X}$ are asymptotically normal under weak assumptions by showing that, given a (non-random) sample $X_1, \dots, X_n$ and a unit vector $\mathbf{w}$ uniformly distributed on the d-dimensional unit sphere, the empirical measure of the sample $\mathbf{w}'\mathbf{X}_1, \dots, \mathbf{w}'\mathbf{X}_n$ converges weakly to a normal law, in probability. Further, the joint distribution of two linear combinations, say, $\mathbf{w}_d'\mathbf{Z}$ and $\mathbf{v}_d'\mathbf{Z}$, of a d-dimensional random vector $\mathbf{Z}$ possessing a Lebesgue density and being standardized, i.e. $E(\mathbf{Z}) = 0$ and $E(\mathbf{Z}\mathbf{Z}') = \mathrm{id}_d$, is bivariate normal in that sense with unit variances and covariance $\mathbf{w}'\mathbf{v}$. However, in that work the projection vectors are *random* and the data are assumed *fixed* (e.g. by conditioning) and constrained to satisfy conditions that are satisfied by, e.g., i.i.d. random vectors of dimension d_n with i.i.d. entries. Contrary, in this paper the projection vectors are fixed and the observations random. Assuming a linear process framework, we provide Gaussian approximations for the sample estimates of the variances of the projections.

3. A FRAMEWORK FOR HIGH-DIMENSIONAL TIME SERIES

3.1. *Model and assumptions*

Let $\{\epsilon_k : k \in \mathbb{Z}\}$ be a sequence of independent random variables with mean zero, variances

$$\sigma_k^2 = E(\epsilon_k^2)$$

and uniformly bounded moments of the order $(4+\delta)$,

$$\sup_k E|\epsilon_k|^{4+\delta} < \infty,$$

for some $\delta > 0$, such that $\gamma_k = E\epsilon_k^4$ and σ_k^2 are finite, for all $k \in \mathbb{Z}$.

We shall assume throughout the paper that the νth coordinate of $\mathbf{Y}_n$ is given by

$$Y_k^{(\nu)} = Y_{nk}^{(\nu)} = \sum_{j=0}^{\infty} c_{nj}^{(\nu)} \epsilon_{k-j}, \qquad k = 1, \ldots, n, \tag{3.1}$$

for coefficients $\{c_{nj}^{(\nu)} : j \in \mathbb{N}_0\}$, $\nu = 1, \ldots, d_n$. We mainly have in mind the case that we observe, at time n, the first n observations of d_n *sequences* $\{Y_k^{(\nu)} : k \geq 0\}$, $\nu = 1, \ldots, d_n$, but our results also allow for *arrays* $\{Y_{nk}^{(\nu)} : k \geq 0, n \geq 1\}$, since the coefficients may depend on n.

Model (3.1) implies that the cross-sectional as well as serial correlations have a specific structure, since

$$\mathrm{Cov}\,(Y_t^{(\nu)}, Y_t^{(\mu)}) = \sum_{j=0}^{\infty} c_{nj}^{(\nu)} c_{nj}^{(\mu)} \sigma_{t-j}^2 \tag{3.2}$$

and for $h > 0$,

$$\mathrm{Cov}\,(Y_{nt}^{(\nu)}, Y_{n,t+h}^{(\mu)}) = \sum_{j=0}^{\infty} c_{nj}^{(\nu)} c_{n,j+h}^{(\mu)} \sigma_{t-j}^2, \tag{3.3}$$

for $1 \leq \nu, \mu \leq d_n$ and all t. Consequently, the cross-sectional variance-covariance matrix $\mathrm{Var}\,(\mathbf{Y}_{nt})$ is given by

$$\mathbf{\Sigma}_n[t] = \mathbf{C}_n \Lambda \mathbf{C}_n' = \sum_{j=0}^{\infty} \sigma_{t-j}^2 \mathbf{c}_{nj} \mathbf{c}_{nj}', \quad \Lambda = \mathrm{diag}(\sigma_0^2, \sigma_1^2, \ldots), \tag{3.4}$$

where $\mathbf{C}_n = (c_{nj}^{(\nu)})_{1 \leq \nu \leq d_n, 1 \leq j}$ is the $(d_n \times \infty)$-dimensional matrix with column vectors

$$\mathbf{c}_{nj} = (c_{nj}^{(1)}, \ldots, c_{nj}^{(d_n)})', \qquad j \geq 0.$$

The lag h serial covariance matrix attains the representation

$$\mathbf{\Sigma}_n(h) = E(\mathbf{Y}_{nt} \mathbf{Y}_{n,t+h}') = \mathbf{C}_n \Lambda (L^{-h} \mathbf{C}_n)'$$

where L denotes the lag operator that acts on all columns, i.e. $L^{-1}\mathbf{C}_n = (c_{n,i+1}^{(\nu)})_{i \geq 0, 1 \leq \nu \leq d_n}$.

We shall impose the following condition on the decay of the coefficients $c_{nj}^{(\nu)}$, which is similar to the assumption imposed in [14], where it controls the principal component eigenvectors within a factor model (also see our discussion below), and to condition (2.4) in [6], where it controls the error terms of a panel time series model.

Assumption (A) The sequences $\{c_{nj}^{(\nu)} : j \in \mathbb{N}_0\}$ satisfy

$$\sup_{n \in \mathbb{N}} \max_{1 \leq \nu \leq d_n} |c_{nj}^{(\nu)}|^2 << j^{-3/2-\theta/2} \tag{3.5}$$

for some $0 < \theta < 1/2$. Here and in the sequel $a_n << b_n$ stands for $a_n = O(b_n)$.

Indeed, (3.5) covers not only short memory processes for which the covariances, say, $r_k = E(X_0 X_k)$, are summable, i.e. $\sum_k |r_k| < \infty$, but also many long-range dependent series. An example for the latter is fractionally integrated noise of order $d \in (-1/2, 1/4 - \theta/2)$, i.e. a stationary solution of the equation

$$(1 - L)^d X_t = \epsilon_t,$$

where $\{\epsilon_t\}$ is a white-noise series, that is given by $X_t = \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}$ with coefficients $\theta_k = \Gamma(k + d)/(\Gamma(k + 1)\Gamma(d)) \sim k^{d-1}/\Gamma(d)$, see e.g. [30].

Notice that Assumption (A) implies that the jth column vector satisfies

$$\mathbf{c}_{nj}^2 = O(\mathbf{1} j^{-3/2-\theta/2}), \tag{3.6}$$

such that $\mathbf{\Sigma}_n$ can be approximated by a relatively small number of eigenvectors by truncating the series (3.4) as shown by the following lemma which estimates the *scaled Frobenius norm*

$$\|\mathbf{A}\|_F^* = \frac{1}{d_n^{1/2}} \left(\sum_{i,j=1}^{d_n} a_{ij}^2 \right)^{1/2}$$

of a $(d_n \times d_n)$-dimensional matrix $\mathbf{A}$.

Lemma 3.1 *Suppose Assumption (A) holds true and that, for fixed t, the variances σ_{t-j}^2, $j \geq 0$, of the innovations satisfy*

$$\sum_{j=0}^{\infty} j^{-3/2-\theta/2} \sigma_{t-j}^2 < \infty.$$

Then

$$\left\| \mathbf{\Sigma}_n[t] - \sum_{j=1}^{r} \sigma_{t-j}^2 \mathbf{c}_{nj} \mathbf{c}_{nj}{}' \right\|_F^* = o(1),$$

as $r \to \infty$.

PROOF: Using (3.4) and the fact that (3.6) yields $\|\mathbf{c}_{nj}\|^2 \leq C d_n j^{-3/2-\theta/2}$ for some constant $C < \infty$, we immediately obtain

$$\left\| \mathbf{\Sigma}_n[t] - \sum_{j=1}^{r} \sigma_{t-j}^2 \mathbf{c}_{nj} \mathbf{c}_{nj}{}' \right\|_F^* \leq \sum_{j>r} \sigma_{t-j}^2 \|\mathbf{c}_{nj}\mathbf{c}_{nj}{}'\|_F^* \leq C \sum_{j>r} \sigma_{t-j}^2 j^{-3/2-\theta/2}.$$

Q.E.D.

It is worth noting that, under certain circumstances, one may interpret model (3.1) as an infinite-dimensional latent one-factor model. For that purpose, let us assume for a moment that $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ is a stationary series, such that $\sigma_k^2 = \sigma^2$ for all k. Then, by stationarity, $\mathbf{Y}_n = (Y^{(1)}, \dots, Y^{(d_n)})'$ satisfies

$$\mathbf{Y}_n \stackrel{d}{=} \mathbf{C}_n \boldsymbol{\epsilon}_n, \quad \boldsymbol{\epsilon}_n = (\epsilon_0, \epsilon_{-1}, \dots), \tag{3.7}$$

$$\mathbf{\Sigma}_n = \sigma^2 \mathbf{C}_n \mathbf{C}_n', \tag{3.8}$$

where $\{\epsilon_n\}$ is the unobservable common factor.

3.2. *Some preparatory approximations*

In the main proof we shall study in detail the linear process $\sum_{j=0}^{\infty} c_j^w \epsilon_{k-j}$, $k \geq 1$, with coefficients

$$c_j^w = \sum_{\nu=1}^{d_n} w_\nu c_j^{(\nu)}, \qquad j \geq 0, \tag{3.9}$$

associated to a weighting vector $\mathbf{w}_n$. Behind our main results are martingale approximations related to that linear process, whose definitions require the following quantities, which are controlled under Assumption (A) by virtue of Lemma and Definition 3.1 given below. Let

$$f_{0,j}^{(n)} = f_{0,j}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = \sum_{\nu,\mu=1}^{d_n} v_\nu w_\mu c_j^{(\nu)} c_j^{(\mu)}, \qquad j = 0, 1, \dots, \tag{3.10}$$

$$f_{l,j}^{(n)} = f_{l,j}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = \sum_{\nu,\mu=1}^{d_n} v_\nu w_\mu [c_j^{(\nu)} c_{j+l}^{(\mu)} + c_j^{(\mu)} c_{j+l}^{(\nu)}], \qquad l = 1, 2, \dots; \ j = 0, 1, \dots, \tag{3.11}$$

and

$$\widetilde{f}_{l,i}^{(n)} = \widetilde{f}_{l,i}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = \sum_{j=i}^{\infty} f_{l,j}^{(n)} = \sum_{j=i}^{\infty} \sum_{\nu,\mu=1}^{d_n} v_\nu w_\mu [c_j^{(\nu)} c_{j+l}^{(\mu)} + c_j^{(\mu)} c_{j+l}^{(\nu)}], \qquad l, i = 0, 1, \dots. \tag{3.12}$$

LEMMA AND DEFINITION 3.1 *Suppose that $\mathbf{v}_n, \mathbf{w}_n$ have uniformly bounded ℓ_1-norm in the sense of equation (2.2). Then Assumption (A) implies*

$$\sup_{n \in \mathbb{N}} \sum_{i=1}^{\infty} \sum_{l=0}^{\infty} (\widetilde{f}_{l,i}^{(n)} - \widetilde{f}_{l,i+n'}^{(n)})^2 \leq C (n')^{1-\theta}, \qquad \textit{for all } n' = 1, 2, \dots, \tag{3.13}$$

$$\sup_{n\in\mathbb{N}}\sum_{k=1}^{n'}\sum_{r=0}^{\infty}(\widetilde{f}_{r+k,0}^{(n)})^2 \le C(n')^{1-\theta}, \qquad \textit{for all } n'=1,2,\dots, \tag{3.14}$$

$$\sup_{n\in\mathbb{N}}\sum_{k=1}^{n'}\sum_{l=0}^{\infty}(\widetilde{f}_{l,k}^{(n)})^2 \le C(n')^{1-\theta}, \qquad \textit{for all } n'=1,2,\dots, \tag{3.15}$$

and there exist

$$\alpha_n^2 = \alpha_n^2(\mathbf{v}_n,\mathbf{w}_n) \ge 0, \qquad n\ge 1, \tag{3.16}$$

such that

$$(\widetilde{f}_{00}^{(n)})^2\sum_{j=1}^{n'}(\gamma_{m'+j}-\sigma_{m'+j}^4)+\sum_{j=1}^{n'}\sum_{l=1}^{j-1}(\widetilde{f}_{j-l,0}^{(n)})^2\sigma_{m'+j}^2\sigma_{m'+l}^2 - n'\alpha_n^2 \le C(n')^{1-\theta}, \tag{3.17}$$

for all $n',m'=0,1,\cdots$.

Further, if $\mathbf{v}_n,\mathbf{w}_n,\widetilde{\mathbf{v}}_n,\widetilde{\mathbf{w}}_n$, $n\ge 1$, *have uniformly bounded* ℓ_1*-norms, then there exist*

$$\beta_n = \beta_n(\mathbf{v}_n,\mathbf{w}_n,\widetilde{\mathbf{v}}_n,\widetilde{\mathbf{w}}_n), \qquad n\ge 1, \tag{3.18}$$

with

$$\widetilde{f}_{0,0}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)\widetilde{f}_{0,0}^{(n)}(\widetilde{\mathbf{v}}_n,\widetilde{\mathbf{w}}_n)\sum_{j=1}^{n'}(\gamma_{m'+j}-\sigma_{m'+j}^4)+\sum_{j=1}^{n'}\sum_{l=1}^{j-1}\widetilde{f}_{j-l,0}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)\widetilde{f}_{j-l,0}^{(n)}(\widetilde{\mathbf{v}}_n,\widetilde{\mathbf{w}}_n)\sigma_{m'+j}^2\sigma_{m'+l}^2 - n'\beta_n(\mathbf{v}_n,\mathbf{w}_n,\widetilde{\mathbf{v}}_n,\widetilde{\mathbf{w}}_n) \overset{n',m'}{<<} (n')^{1-\theta}. \tag{3.19}$$

PROOF: This follows from

$$\sup_{n\in\mathbb{N}}(c_j^w)^2=\sup_{n\in\mathbb{N}}\left(\sum_{\nu=1}^{d_n}w_\nu c_j^{(\nu)}\right)^2 \le \sup_{n\in\mathbb{N}}\max_{1\le j\le d_n}|c_{nj}^{\nu}|^2\|\mathbf{w}_n\|_{\ell_1}^2$$

and [18, Remark 3.2]. *Q.E.D.*

REMARK 3.1 *If the moments up to the order* 4 *are stationary such that* $\gamma_k=\gamma$ *and* $\sigma_k^2=\sigma^2$, *say, (3.17) is a consequence of (3.14) and*

$$\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n)=(\gamma-\sigma^4)[\widetilde{f}_{0,0}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)]^2+\sigma^4\sum_{l=1}^{\infty}[\widetilde{f}_{l,0}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)]^2,$$

as well as

$$\beta_n(\mathbf{v}_n,\mathbf{w}_n,\widetilde{\mathbf{v}}_n,\widetilde{\mathbf{w}}_n)=\widetilde{f}_{0,0}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)\widetilde{f}_{0,0}^{(n)}(\widetilde{\mathbf{v}}_n,\widetilde{\mathbf{w}}_n)(\gamma-\sigma^4)+\sigma^4\sum_{l=1}^{\infty}\widetilde{f}_{l,0}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)\widetilde{f}_{l,0}^{(n)}(\widetilde{\mathbf{v}}_n,\widetilde{\mathbf{w}}_n).$$

Note also that, in general, (3.13)–(3.15) and (3.17) ensure that $\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n)$ *and* $\beta_n(\mathbf{v}_n,\mathbf{w}_n,\widetilde{\mathbf{v}}_n,\widetilde{\mathbf{w}}_n)$ *are bounded.*

4. ASYMPTOTICS

For many sequences, X_n, $n \geq 1$, of random variables, such as sums of i.i.d. random variables with finite second moment, a strong approximation with rate holds true, i.e. after redefining the process on a new probability space, there exists a Brownian motion $B(t)$, $t \geq 0$, such that, on that new space,

$$|X_t - \sigma B(t)| = O(t^{1/2-\lambda}), \qquad \text{for all } t > 0,$$

a.s., as $n \to \infty$, for some positive constant σ and a constant $\lambda > 0$. Results of this type date back to the seminal work of [16], [17] and have been extended and refined since then, in particular to martingales, see e.g. [22]. Such a strong approximation result also yields an approximation of the rescaled càdlàg process $n^{-1/2}X_{\lfloor tn \rfloor}$, $t \in [0,1]$, by the Brownian motion $\sigma B(t)$, $t \in [0,1]$, and implies the FCLT, i.e. the weak convergence

$$n^{-1/2}X_{\lfloor tn \rfloor} \Rightarrow \sigma B(t),$$

as $n \to \infty$, where $\Rightarrow$ signifies weak convergence in the Skorohod space $D[0,1]$. This, in turn, implies the weak convergence of continuous mappings of $n^{-1/2}X_n(\lfloor n\bullet \rfloor)$. Observe that when $X_n = \sum_{i=1}^n \xi_i$ is a sum of i.i.d. random variables with $E(\xi_1^2) < \infty$, then we obtain the classical Donsker theorem and, for $t = 1$, the CLT.

In order to obtain strong approximations for the bilinear form $Q(\mathbf{v}_n, \mathbf{w}_n)$, we shall derive a martingale approximation for a partial sum associated to $Q(\mathbf{v}_n, \mathbf{w}_n)$, to which conditions due to [22] for the validity of a strong approximation can be applied.

We need to introduce further notation. Define

$$\widehat{\mathbf{\Sigma}}_{nk} = \left(\sum_{i=1}^k Y_i^{(\nu)} Y_i^{(\mu)} \right)_{1 \leq \nu,\mu \leq d_n}, \tag{4.1}$$

$$\mathbf{\Sigma}_{nk} = \left(\sum_{i=1}^k E Y_i^{(\nu)} Y_i^{(\mu)} \right)_{1 \leq \nu,\mu \leq d_n}, \tag{4.2}$$

for $n, k \geq 1$. To be precise, our results shall deal with

$$D_{nk} = \mathbf{v}_n'(\widehat{\mathbf{\Sigma}}_{nk} - \mathbf{\Sigma}_{nk})\mathbf{w}_n, \qquad n, k \geq 1,$$

and the associated càdlàg processes

$$\mathcal{D}_n(t) = \mathbf{v}_n' n^{-1/2}(\widehat{\mathbf{\Sigma}}_{n,\lfloor nt \rfloor} - \mathbf{\Sigma}_{n,\lfloor nt \rfloor})\mathbf{w}_n, \qquad t \in [0,1], n \geq 1.$$

Notice that

$$\mathcal{D}_n(1) = \mathbf{v}_n' \sqrt{n}(\widehat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n)\mathbf{w}_n, \qquad n \geq 1,$$

is the centered and scaled version of the bilinear form $Q(\mathbf{v}_n, \mathbf{w}_n)$, where

$$\boldsymbol{\Sigma}_n = E\widehat{\boldsymbol{\Sigma}}_n = \frac{1}{n}\sum_{i=1}^{n} E(\mathbf{Y}_{ni}\mathbf{Y}_{ni})',$$

If $\{\mathbf{Y}_{ni} : 1 \le i \le n\}$ is stationary, then $\boldsymbol{\Sigma}_n$ simplifies to $\boldsymbol{\Sigma}_n = E(\mathbf{Y}_{n1}\mathbf{Y}_{n1}')$.

If the dependence of the above quantities on $\mathbf{v}_n, \mathbf{w}_n$ matters, we shall indicate this in our notation and then write

$$D_{nk}(\mathbf{v}_n, \mathbf{w}_n), \mathcal{D}_n(t; \mathbf{v}_n, \mathbf{w}_n), \mathcal{D}_n(1; \mathbf{v}_n, \mathbf{w}_n).$$

For two weighting vectors $\mathbf{v}_n$ and $\mathbf{w}_n$ we may associate the martingales

$$M_k^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = \widetilde{f}_{00}^{(n)}(\mathbf{v}_n, \mathbf{w}_n)\sum_{i=0}^{k}(\epsilon_i^2 - \sigma_i^2) + \sum_{i=0}^{k}\epsilon_i \sum_{l=0}^{\infty}\widetilde{f}_{l,0}^{(n)}(\mathbf{v}_n, \mathbf{w}_n)\epsilon_{i-l}, \qquad k, n \ge 0, \tag{4.3}$$

and the associated processes

$$\mathcal{M}_n(t; \mathbf{v}_n, \mathbf{w}_n) = n^{-1/2} M_{\lfloor nt \rfloor}^{(n)}(\mathbf{v}_n, \mathbf{w}_n). \qquad t \in [0,1], n \ge 1,$$

It turns out that, under the assumptions of the article, those martingales are close to $D_{nk}(\mathbf{v}_n, \mathbf{w}_n)$ and $\mathcal{D}_n(t; \mathbf{v}_n, \mathbf{w}_n)$, respectively, which is the key to obtain large-sample asymptotics in terms of strong approximations and second order information, i.e. variances and covariances.

The following theorem shows that bilinear forms of uniformly bounded ℓ_1-projections satisfy a strong invariance principle with rate.

Theorem 4.1 *Suppose Assumption (A) holds true. Then, for each $n \in \mathbb{N}$, there exists an equivalent version of $D_n(t)$, $t \ge 0$, again denoted by $D_n(t)$, and a standard Brownian motion $\{B_n(t) : t \ge 0\}$, which depends on $(\mathbf{v}_n, \mathbf{w}_n)$, i.e. $B_n(t) = B_n(t; \mathbf{v}_n, \mathbf{w}_n)$, both defined on some probability space $(\Omega_n, \mathcal{F}_n, P_n)$, such that for some $\lambda > 0$*

$$|D_{nt} - \alpha_n B_n(t)| \le C t^{1/2-\lambda}, \qquad \textit{for all } t > 0 \textit{ a.s.},$$

as $n \to \infty$. This implies the strong approximation

$$\sup_{t\in[0,1]} |\mathcal{D}_n(t) - \alpha_n B_n(\lfloor nt \rfloor / n)| \le C n^{-\lambda}, \qquad \textit{a.s.},$$

as $n \to \infty$, as well as the CLT

$$|\mathcal{D}_n(1) - \alpha_n B_n(1)| = o(1), \qquad \textit{a.s.},$$

as $n \to \infty$, i.e. $\mathcal{D}_n(1)$ is asymptotically $\mathcal{N}(0, \alpha_n^2)$.

PROOF: Notice that we have

$$\begin{aligned} D_{nk}(\mathbf{v}_n, \mathbf{w}_n) &= \mathbf{v}_n'(\widehat{\mathbf{\Sigma}}_{n,k} - \mathbf{\Sigma}_{n,k})\mathbf{w}_n \\ &= \sum_{\nu,\mu=1}^{d_n} v_\nu w_\mu \sum_{i \le k} [Y_i^{(\nu)} Y_i^{(\mu)} - E Y_i^{(\nu)} Y_i^{(\mu)}] \\ &= \sum_{i \le k} \left\{ \sum_{\nu,\mu=1}^{d_n} v_\nu w_\mu Y_i^{(\nu)} Y_i^{(\mu)} - \sum_{\nu,\mu=1}^{d_n} v_\nu w_\mu E Y_i^{(\nu)} Y_i^{(\mu)} \right\} \end{aligned}$$

leading to the representation

$$D_{nk}(\mathbf{v}_n, \mathbf{w}_n) = \sum_{i \le k} [Y_{ni}(\mathbf{v}_n) Y_{ni}(\mathbf{w}_n) - E Y_{ni}(\mathbf{v}_n) Y_{ni}(\mathbf{w}_n)]$$

with linear processes

$$\begin{aligned} Y_{ni}(\mathbf{v}_n) &= \sum_{j=0}^{\infty} c_{nj}^{(v)} \epsilon_{i-j}, \\ Y_{ni}(\mathbf{w}_n) &= \sum_{j=0}^{\infty} c_{nj}^{(w)} \epsilon_{i-j} \end{aligned}$$

w.r.t. $\{\epsilon_t\}$ given by the coefficients

$$c_{nj}^{(v)} = \sum_{\nu=1}^{d_n} v_\nu c_{nj}^{(\nu)}, \qquad c_{nj}^{(w)} = \sum_{\nu=1}^{d_n} w_\nu c_{nj}^{(\nu)},$$

for $j \ge 0$ and $n \ge 1$. We may now follow the method of proof of [18], however, we have to take into account that the above processes depend on n.

Let $\mathcal{F}_m = \sigma(\epsilon_i : i \le m)$, $m \ge 1$. It is easy to check that, for any fixed $n \in \mathbb{N}$, the r.v.s.

$$M_m^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = \widetilde{f}_{0,0}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) \sum_{k=0}^{m} (\epsilon_k^2 - \sigma_k^2) + \sum_{k=0}^{m} \epsilon_k \sum_{l=0}^{\infty} \widetilde{f}_{l,0}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) \epsilon_{k-l}, \qquad m \ge 0,$$

satisfy $E(M_m^{(n)}(\mathbf{v}_n, \mathbf{w}_n) \,|\, \mathcal{F}_{m-1}) = M_{m-1}^{(n)}(\mathbf{v}_n, \mathbf{w}_n)$, for $m \ge 0$, thus forming a martingale array $\{M_m^{(n)}(\mathbf{v}_n, \mathbf{w}_n) : m \in \mathbb{N}, n \in \mathbb{N}\}$ with associated martingale differences

$$M_{n'+m'}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) - M_{m'}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = \widetilde{f}_{0,0}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) \sum_{k=m'+1}^{n'+m'} (\epsilon_k^2 - \sigma_k^2) + \sum_{k=m'+1}^{n'+m'} \epsilon_k \sum_{l=0}^{\infty} \widetilde{f}_{l,0}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) \epsilon_{k-l},$$

for $n', m' \ge 0$. Put

$$D_{n',m'}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = \sum_{k=m'+1}^{m'+n'} [Y_k(\mathbf{v}_n) Y_k(\mathbf{w}_n) - E Y_k(\mathbf{v}_n) Y_k(\mathbf{w}_n)], \qquad m', n' \ge 0,$$

and consider the decomposition

$$D_{n',m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n) = M_{n'+m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n) - M_{m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n) + R_{n',m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n), \qquad m',n' \geq 0.$$

In order to justify the approximation of $D_{n',m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)$ by the martingale differences defined above, it suffices to show that $\sup_n E[R_{n',m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)]^2$ tends to 0 sufficiently fast, as $n',m' \to \infty$. Using the representation [18, (4.3)], repeating the arguments in [18] leading to the bounds in (4.8), (4.9) and (4.10) therein and noting that those bounds are uniform in $n \geq 1$, we obtain

$$E(R_{n'm'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n))^2 \overset{n',m'}{<<} (n')^{-3/2-\theta}.$$

and, for each $n \in \mathbb{N}$,

$$\|E[(D_{m'n'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n))^2 \,|\, \mathcal{F}_{m'}]\|_1 << (n')^{-3/2-\theta}.$$

Here $a_{nm} \overset{n,m}{<<} b_{nm}$ means that there is a constant c such that $|a_{nm}| \leq c|b_{nm}|$ for all n,m. This approximation in L_2 with a rate allows for very general conditions for the validity of a strong approximation. Again, we may follow the arguments given by [18], by verifying the following sufficient conditions due to [22]. In terms of an array $\xi_k^{(n)}$, $k=1,\dots,n$, of r.v.s., Philipp's result is as follows. Let $\mathcal{G}_m^{(n)} = \sigma(\xi_{ni} : i \leq m)$. If

$$S_{n',m'}^{(n)} = \sum_{k=m'+1}^{m'+n'} \xi_k^{(n)}, m', n' \geq 0, \tag{4.4}$$

satisfies

(I) $\|E(S_{n',m'}^{(n)}|\mathcal{G}_{m'}^{(n)})\|_1 \overset{m',n'}{<<} (n')^{1/2-\varepsilon}$, a.s., for some $\varepsilon > 0$,

(II) there exists an $\alpha_n^2 \geq 0$ such that $\|E[(S_{n',m'}^{(n)})^2|\mathcal{G}_{m'}^{(n)}] - n'\alpha_n^2\|_1 \overset{n',m'}{<<} (n')^{1-\varepsilon}$, a.s., for some $\varepsilon > 0$.

(III) $\sup_{k\geq 0} E|\xi_k^{(n)}|^{4+\delta} < \infty$ for some $\delta > 0$,

then there exists a process $\{\widetilde{S}_{n'}^{(n)} : n' \geq 0\}$ and a standard Brownian motion $\{\widetilde{B}_t^{(n)} : t \geq 0\}$ on some probability space $(\widetilde{\Omega},\widetilde{\mathcal{F}},\widetilde{P})$, such that $\{\widetilde{S}_{n'}^{(n)} : n' \geq 0\} \overset{d}{=} \{S_{n',0}^{(n)} : n' \geq 0\}$ and for some $\lambda > 0$

$$|\widetilde{S}_{\lfloor t \rfloor}^{(n)} - \alpha_n \widetilde{B}_t^{(n)}| \overset{t}{<<} t^{1/2-\lambda},$$

for all $t > 0$ $\widetilde{P}$-a.s. Putting, for fixed $n \geq 1$,

$$\xi_k^{(n)} = Y_k(\mathbf{v}_n)Y_k(\mathbf{w}_n) - E(Y_k(\mathbf{v}_n)Y_k(\mathbf{w}_n)), \tag{4.5}$$

$\mathcal{G}_m^{(n)} = \mathcal{F}_m$ (since the $\xi_k^{(n)}$ are $\mathcal{F}_k$–measurable) and repeating the arguments of [18], we see that, by virtue of Assumption (A), (I)-(III) hold true, which establishes the existence of $B_n(t)$ such that

$$|D_{nt} - \alpha_n B_n(t)| \leq Ct^{1/2-\lambda},$$

for all $t > 0$, a.s. By virtue of the scaling property of Brownian motion, this also implies

$$\sup_{t\in[0,1]} |n^{-1/2}D_{n,\lfloor nt\rfloor} - \alpha_n B_n(\lfloor nt\rfloor/n)| << n^{-\lambda}.$$

It also follows that, for each fixed n, the conditional variance of $M_{m'+n'}^{(n)} - M_{m'}^{(n)}$ satisfies

$$\|E[(M_{m'+n'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n) - M_{m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n))^2|\mathcal{F}_{m'}] - n'\alpha_n^2\|_1 \stackrel{n',m'}{<<} (n')^{1-\theta/2} \tag{4.6}$$

and (cf. [18, (4.22)])

$$\|E[(D_{n',m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n))^2 \,|\, \mathcal{F}_{m'}] - n'\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n)\|_1 \stackrel{n',m'}{<<} (n')^{1-\theta/2} \tag{4.7}$$

as well as

$$|E(D_{n',m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)^2 - n'\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n)| \stackrel{n',m'}{<<} (n')^{1-\theta/2}.$$

Q.E.D.

The extension to a finite number, K, of bilinear forms is straightforward. In particular, we have the following

Corollary 4.1 *Let $\{\mathbf{v}_{nj},\mathbf{w}_{nj} : 1 \le j \le K\}$ be weighting vectors of dimension d_n satisfying condition (2.2). Then, under the assumptions of Theorem 4.1, there exists a K–dimensional Brownian motion with coordinates $B_n(t;\mathbf{v}_{ni},\mathbf{w}_{ni})$ satisfying*

$$E(B_n(t;\mathbf{v}_{ni},\mathbf{w}_{ni})B_n(t;\mathbf{v}_{nj},\mathbf{w}_{nj})) = \beta_n(\mathbf{v}_{ni},\mathbf{w}_{ni},\mathbf{v}_{nj},\mathbf{w}_{nj}),$$

for $1 \le i,j \le K$, such that

$$\left\|(\mathcal{D}_n(t;\mathbf{v}_{ni},\mathbf{w}_{ni})_{i=1}^K - (B_n(\lfloor nt\rfloor/n;\mathbf{v}_{ni},\mathbf{w}_{ni})_{i=1}^K\right\| = O(n^{-\lambda}),$$

a.s., for some constant $\lambda > 0$.

Having in mind the application of the above result to change-point detection, the following corollary dealing with the frequently used maximally selected CUSUM statistic is of interest.

Corollary 4.2 *Suppose that $\mathbf{Y}_{n1},\dots,\mathbf{Y}_{nn}$ is a d_n–dimensional vector time series satisfying Assumption (A). Then, after redefining the series on a new probability space, there exists a Brownian motion such that*

$$\left|\max_{k\le n} |\mathcal{D}_n(k/n)| - \max_{k\le n} |\alpha_n B_n(k/n)|\right| \le Cn^{-\lambda}$$

a.s. for some constant $\lambda > 0$.

5. APPLICATIONS

The results of the present paper have direct applications to several problems and procedures, respectively, which are extensively studied for high–dimensional time series. They contribute novel large-sample approximations for making inference based on the corresponding statistics, usually projections.

5.1. *Optimal portfolio selection*

The problem of optimal portfolio selection, dating back to Markowitz' seminal work, [21] is an intrinsically high-dimensional problem. Given a usually large number d_n of assets with time t returns $R_t^{(j)}$, $j = 1, \dots, d_n$, with mean vector μ and covariance matrix $\mathbf{\Sigma}_n$ of the vector $\mathbf{R}_t = (R_t^{(1)}, \dots, R_n^{d_n})'$ of asset returns, a portfolio is a vector $\mathbf{w}_n = (w_{n1}, \dots, w_{nd_n})'$ with $\sum_{j=1}^{d_n} w_{nj} = 1$, i.e. w_{nj} is the percentage of the capital that is invested into share j. A classical formulation is to minimize the risk, defined as the variance, associated to the portfolio return $\mathbf{w}_n'\mathbf{R}_n$ at time n, i.e. to consider the problem

$$\min_{\mathbf{w}_n} \operatorname{Var}(\mathbf{w}_n'\mathbf{R}_n) = \mathbf{w}_n'\mathbf{\Sigma}_n\mathbf{w}_n, \qquad \text{subject to } \mathbf{w}_n'\mathbf{1} = 1,$$

whose solution is known to be

$$\mathbf{w}_n = (\mathbf{1}'\Sigma_n^{-1}\mathbf{1})^{-1}\mathbf{1}'\Sigma_n^{-1}$$

where $\mathbf{1}$ is the n-vector with unit entries. Obviously, if $\mathbf{w}_n \geq 0$ (no short sales), then

$$\|\mathbf{w}_n\|_{\ell_1} = \mathbf{w}_n'\mathbf{1} = (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}\mathbf{1}'\Sigma_n^{-1}\mathbf{1} = 1,$$

such that the optimal portfolio has uniformly bounded ℓ_1-norm. The mean-variance formulation adds a constraint on the target mean portfolio return and thus considers the problem

$$\min_{\mathbf{w}_n} \mathbf{w}_n'\mathbf{\Sigma}_n\mathbf{w}_n, \qquad \text{subject to } \mathbf{w}_n'\mathbf{1} = 1, \; \mathbf{w}_n'\mu = \mu_0$$

for some μ_0. The solution is

$$\mathbf{w}_n = \frac{c - \mu_0 b}{ac - b^2}\mathbf{\Sigma}_n^{-1}\mathbf{1} + \frac{\mu_0 a - b}{ac - b^2}\mathbf{\Sigma}_n^{-1}\mu.$$

with $a = \mathbf{1}'\mathbf{\Sigma}_n^{-1}\mathbf{1}$, $b = \mathbf{1}'\mathbf{\Sigma}_n^{-1}\mu$ and $c = \mu'\mathbf{\Sigma}_n^{-1}\mu$. Based on estimates of $\mathbf{\Sigma}_n^{-1}$ and μ from past data, one calculates the optimal portfolio which is then held until the next rebalancing. If the dimension is large, inverting the sample covariance matrix may result in substantial numerical instability and becomes impossible if the dimension is larger than the sample size. Shrinking is a commonly applied approach for regularization, in order to obtain a stable and invertible estimator, see [19], amongst others. (We also refer to our subsection 5.3 for more on this.) It is worth mentioning that adding a non-negativity constraint (i.e. no-short-sales) has been observed to have a similar regularizing effect, see [13].To obtain sparse portfolios, [5] proposed to add an ℓ_1-constraint. Observing that Markowitz' problem is equivalent to

$$\min_{\mathbf{w}_n} E(|\mu_0 - \mathbf{w}_n'\mathbf{R}_n|^2), \qquad \text{subject to } \mathbf{w}_n'\mu = \mu_0,\ \mathbf{w}_n'\mathbf{1} = 1,$$

which suggests the empirical version

$$\min_{\mathbf{w}_n} n^{-1}\|\mu_0\mathbf{1} - \mathcal{X}_n\mathbf{w}_n\|_{\ell_2}^2, \qquad \text{subject to } \mathbf{w}_n'\widehat{\mu}_n = \mu_0,\ \mathbf{w}_n'\mathbf{1} = 1,$$

where $\mathcal{X}_n = (R_t^{(j)})_{t,j}$ is the $n \times d_n$ dimensional data matrix of returns and $\widehat{\mu}_n = n^{-1}\sum_{t=1}^n \mathbf{R}_t$, see [5, formula 1], they examined the ℓ_1-regularized version given by

$$\min_{\mathbf{w}_n} n^{-1}\|\mu_0\mathbf{1} - \mathcal{X}_n\mathbf{w}_n\|_{\ell_2}^2 + \lambda\|\mathbf{w}_n\|_{\ell_1}$$

subject to

$$\mathbf{w}_n'\widehat{\mu}_n = \mu_0,\ \mathbf{w}_n'\mathbf{1} = 1,$$

for some regularization parameter $\lambda > 0$. Whereas for large values of λ the classical solution is recovered, smaller values lead to effective ℓ_1-penalization and sparse portfolio vectors with only relatively few active positions.

5.2. *Projections onto lower-dimensional subspaces and sparse principal components*

A primary goal of multivariate statistical analysis is to project multivariate data onto lower-dimensional subspaces, and thus the results of this paper directly address various approaches that are based on ℓ_1 projection vectors. In particular, the results apply to the recently proposed methods of [15], [24] and [35] for sparse principal component analysis.

First observe that the linear combination, $\mathbf{w}_n'\mathbf{Y}_n$, is related to the one-dimensional projection onto the subspace span$\{\mathbf{w}_n\}$, given by

$$\frac{\mathbf{w}_n'\mathbf{Y}_n}{\|\mathbf{w}_n\|_{\ell_2}^2}\mathbf{w}_n.$$

More generally, linear mappings onto a lower-dimensional subspace spanned by such vectors are frequently studied,

$$\pi_n = \mathbf{P}_n\mathbf{P}_n'\mathbf{Y}_n, \qquad \mathbf{P}_n = [\mathbf{w}_n^{(1)}, \dots, \mathbf{w}_n^{(L)}],$$

where $\mathbf{w}_n^{(k)}$, $k = 1, \dots, L$, are, say, q-dimensional vectors, $q < p$. Notice that π_n is an orthogonal projection, if the vectors are orthonormal. An ℓ_1-constraint

$$\|\mathbf{w}_n^{(j)}\|_{\ell_1} \le c, \qquad j = 1, \dots, L,$$

for some constant c, can be ensured by solving the ℓ_1-constrained optimal projection onto $\operatorname{span}\{\mathbf{w}_n^{(j)}\}$,

$$\max_{\mathbf{u}} \mathbf{u}'\mathbf{w}_n^{(j)} \qquad \text{subject to } \|\mathbf{u}\|_{\ell_2}^2 \le 1, \|\mathbf{u}\|_{\ell_1} \le c,$$

whose solution is

$$\widetilde{\mathbf{w}}_n^{(j)} = \frac{S(\mathbf{w}_n^{(j)}, \delta)}{\|S(\mathbf{w}_n^{(j)}, \delta)\|_{\ell_2}},$$

where

$$S(a, \delta) = \operatorname{sgn}(a)(|a| - \delta)_+$$

is the soft-thresholding function, $x_+ = x$ if $x > 0$ and $= 0$ otherwise, and $\delta \ge 0$ is chosen such that $\|S(\mathbf{w}_n^{(j)}, \delta)\|_{\ell_1} = c$, see [35, Lemma 2.2] and [24]. Then the orthogonal projection is given by $\mathbf{P}_n(\mathbf{P}_n'\mathbf{P}_n)^{-}\mathbf{P}_n'$, where $\mathbf{A}^{-}$ denotes a generalized inverse of a square matrix $\mathbf{A}$. In any case, $\mathbf{P}_n'\mathbf{Y}_n$ is the dimension-reducing statistic, and inferential procedures for π_n can be based on the asymptotics of $\mathbf{P}_n'\mathbf{Y}_n$, such that our results apply.

The definition of lower-dimensional subspaces generated by ℓ_1-vectors is extensively studied in the field of sparse principal component analysis for high-dimensional data. Let $\mathcal{X}_n$ be a $n \times d_n$-dimensional data matrix with centered columns corresponding to an independent sample of size n of the variables $Y^{(1)}, \dots, Y^{(d_n)}$, whose columns are assumed to be centered. The simplified component technique-lasso (SCoTLASS) approach of [15] defines the first sparse principal component as a solution of the optimization problem

$$\max_{\mathbf{v}} \mathbf{v}'\mathcal{X}_n'\mathcal{X}_n\mathbf{v}, \qquad \text{subject to } \|\mathbf{v}\|_{\ell_2}^2 \le 1,\ \|\mathbf{v}\|_{\ell_1} \le c.$$

Further sparse components are obtained by maximizing the same objective function under above constraints and the additional constraints that the further component is orthogonal to the previous components. In this way, after L steps we obtain L orthogonal ℓ_1-vectors.

In a similar way, [35] propose a sparse principal component analysis by solving, for the first component, the penalized matrix decomposition problem (PMD) with ℓ_1-constraints,

$$\max_{\mathbf{u},\mathbf{v}} \mathbf{u}'\mathcal{X}_n'\mathcal{X}\mathbf{v}, \qquad \text{subject to } \|\mathbf{v}\|_{\ell_1} \le c,\ \|\mathbf{u}\|_{\ell_2}^2 \le 1, \|\mathbf{v}\|_{\ell_2}^2 \le 1,$$

Observing that for fixed $\mathbf{v}$ the solution is given by $\mathbf{u} = \mathcal{X}_n\mathbf{v}/\|\mathcal{X}_n\mathbf{v}\|_{\ell_2}$, the first sparse principal component of PMD with ℓ_1-constraints also solves SCoTLASS, see [35, p. 525]. However, the PMD approach does not constrain the further components to be orthogonal, so that it differs from SCoTLASS. Closely related is the sparse PCA (SPCA) method of [24]. They consider the problem to determine a regularized low-rank matrix approximation,

$$\min_{\mathbf{u},\mathbf{v}} \|\mathcal{X} - \mathbf{u}\mathbf{v}'\|_F^2 + p_\lambda(\mathbf{v}), \qquad \|\mathbf{u}\|_{\ell_2} = 1,$$

for several penalty terms $p_\lambda(\mathbf{v})$ including the case $\lambda\|\mathbf{v}\|_{\ell_1}$.

The lassoed principal components (LPC) method of [34] regresses a n–dimensional score statistic $\mathbf{T}_n$ onto the right eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ of a singular value decomposition of the sample covariance matrix $\mathcal{X}_n$,

$$\mathcal{X}_n = \mathbf{U}_n \mathbf{D}_n \mathbf{V}_n', \qquad \mathbf{V}_n = (\mathbf{v}_1, \dots, \mathbf{v}_n),$$

with a ℓ_1-penality term, known as the lasso, see [32] and [33]. Thus, one considers the problem

$$\min_{\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^n} \|\mathbf{T}_n - \beta_0 - \mathbf{V}_n \beta\|_{\ell_2}^2 + \lambda \|\beta\|_{\ell_1},$$

which is equivalent to

$$\min_{\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^n} \|\mathbf{T}_n - \beta_0 - \mathbf{V}_n \beta\|_{\ell_2}^2, \qquad \text{subject to } \|\beta\|_{\ell_1} \le \delta,$$

such that the resulting estimated regression vector is δ-sparse.

5.3. *Shrinkage estimation*

In many applications, from a statistical point of view, when estimating the common variance-covariance matrix $\mathbf{\Sigma}_n$ of a stationary vector time series $\mathbf{Y}_{n1}, \dots, \mathbf{Y}_{nn}$ of dimension d_n, one has an interest to regularise $\widehat{\mathbf{\Sigma}}_n$ to improve its (finite-sample) properties such as its mean-squared error $E[\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_n\|_F^2]$ or its condition number, defined to be the ratio of its largest to its smallest eigenvalue. This is of particular interest if one needs an invertible estimator of $\mathbf{\Sigma}_n$. One well-established possibility to regularise $\widehat{\mathbf{\Sigma}}_n$ ([20], [23]) is to consider a shrinkage estimator defined by a linear (in fact a convex) combination of $\widehat{\mathbf{\Sigma}}_n$ with a well-conditioned "target". Already in the population, for situations where the dimensionality d_n is in the order magnitude of the sample size n, shrinkage of the high-dimensional variance-covariance matrix $\mathbf{\Sigma}_n$ towards a target, similarly to ridge regression, can reduce a potentially large condition number. This is achieved by reducing the dispersion of the eigenvalues of $\mathbf{\Sigma}_n$ around its "grand mean" $\mu_n := d_n^{-1} \operatorname{tr} \mathbf{\Sigma}_n$: large eigenvalues are pulled down towards μ_n, small eigenvalues are lifted up to μ_n. Improvement of the mean-squared error $E[\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_n\|_F^2]$ is achieved via a potentially tremendous variance reduction (due to stabilisation via regularisation), even if obviously a bias is introduced by adding a deliberately misspecified shrinkage target of low complexity (but high regularity) which usually underfits the true underlying variance-covariance matrix.

A comparatively straightforward, but in practice often well working, choice of the target is a multiple of the d−dimensional identity matrix I_n ([20], [10]). Other choices consist in specifying, e.g. in a context of an economic time series panel, a given or a latent factor which describes the "mean-behaviour" of the panel well in terms of a low-dimensional and hence very stable approximation to the high-dimensional panel structure ([19], [3]). Similarly, adding a parametric estimator of small complexity to the fully nonparametric sample estimator as in [11], follows the same aforementioned paradigm of reducing variance by adding a (model) bias.

The success of this approach, quite naturally, lies in the correct specification of the shrinkage weight W_n, the proportion with which the shrinkage target enters into the convex combination: obviously it has to be the higher the less regular the given variance-covariance matrix. More specifically, in the above mentioned literature, a theory of optimal choice of W_n has been delivered, for various scenario, by minimising the mean squared error between the shrunken estimator $\mathbf{\Sigma}_n^s$ and the true variance-covariance matrix $\mathbf{\Sigma}_n$. Hence, let

$$\mathbf{\Sigma}_n^s = \mathbf{\Sigma}_n^s(W_n) = (1 - W_n)\widehat{\mathbf{\Sigma}}_n \ + \ W_n \ \mu_n I_n,$$

which shrinks the sample covariance matrix towards the shrinkage target $\mu_n I_n$. In the population, the optimal shrinkage weight is derived as

$$W_n^* = \mathrm{argmin}_{W_n \in [0,1]} d_n^{-1} E[\|\mathbf{\Sigma}_n^s(W_n) - \mathbf{\Sigma}_n\|_F^2] \ ,$$

leading to the MSE-optimally shrunken matrix $\mathbf{\Sigma}_n^* = \mathbf{\Sigma}_n^s(W_n^*)$. A closed form solution can be derived as

$$W_n^* = \frac{E[\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_n\|_F^2]}{E[\|\mu_n I_n - \widehat{\mathbf{\Sigma}}_n\|_F^2]} \ ,$$

where one observes the trade-off between the distance of $\widehat{\mathbf{\Sigma}}_n$ towards $\mathbf{\Sigma}_n$ (being large in case of a badly conditioned sample covariance matrix) and the distance of the sample estimator to the shrinkage target. This choice leads to a true improvement on the level of the mean-squared error:

$$E[\|\mathbf{\Sigma}_n^* - \mathbf{\Sigma}_n\|_F^2] < E[\|\widehat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n\|_F^2] \ .$$

It is obvious that our distributional results of Section 4 can be directly applied to the shrunken matrix $\mathbf{\Sigma}_n^s$, this being the first step in the direction of some inference theory for this kind of shrinkage estimators (which is still lacking in the literature). In practice, the population quantities μ_n and W_n^* need to be replaced by some estimators. In the situation of disposing of independent copies of the sampled data, a possibility is to use those, quite analogously to Section 5.2 and many other "statistical learning situations", in order to construct these estimators $\widehat{\mu}_n$ and $\widehat{W}_n$. Then, the discussed results on inference on $\mathbf{\Sigma}_n^*$ continue to hold (conditionally on the "learning sample").

5.4. *Change-point analysis*

Change-point analysis is concerned with the detection and analysis of possible structural changes in the distribution of observations and the determination of the time points of their occurence called change-points. For general methodological overviews we refer to [8], [30] and the references given therein, amongst others. In view of Theorem 4.1 and Corollary 4.2, the following a posteriori (off-line) procedure can be used to test for the presence of a change of the high-dimensional covariance structure having observed the series up to current time n.

Suppose that under the null hypothesis of no change $\mathbf{Y}_{n1}, \dots, \mathbf{Y}_{nn}$ forms a d_n–dimensional mean zero stationary vector time series with variance-covariance matrix $\mathbf{\Sigma}_n^{(0)}$ and satisfying the assumptions of Corollary 4.2. Let us further assume that under the change-point alternative hypothesis the sequence

$$\mathbf{\Sigma}_n[i] = \operatorname{Cov}(\mathbf{Y}_{ni}), \qquad 1 \le i \le n,$$

of variance-covariance matrices is equal to $\mathbf{\Sigma}_n^{(0)}$ up to the change-point $q \in \{1, \dots, n-1\}$ and changes for $i > q$ in such a way that the functional

$$\sigma_n^2(i) = \mathbf{w}_n' \mathbf{\Sigma}_n[i] \mathbf{w}_n,$$

where $\mathbf{w}_n$ satisfies the uniform ℓ_1–condition (2.2), changes from $\sigma_{n0}^2 = \mathbf{w}_n' \mathbf{\Sigma}_n^{(0)} \mathbf{w}_n$ to some different value $\sigma_{n1}^2 \neq \sigma_{n0}^2$ and then remains constant again. An appropriate change-point test statistic directly suggested by our results is given by

$$V_n = \max_{k \le n} |\alpha_n^{-1} \mathcal{D}_n(k/n)| = \max_{k \le n} n^{-1/2} \alpha_n^{-1} |\mathbf{w}_n' (\widehat{\mathbf{\Sigma}}_{nk} - \mathbf{\Sigma}_{nk}) \mathbf{w}_n|$$

where $\widehat{\mathbf{\Sigma}}_{ni}$ and $\mathbf{\Sigma}_{ni}$ are defined in (4.1) and (4.2). Corollary 4.2 now provides us with the asymptotic null distribution,

$$\max_{k \le n} |\alpha_n^{-1} \mathcal{D}_n(k/n)| \sim_{n \to \infty} \sup_{t \in [0,1]} |B(t)|,$$

that is needed to determine critical values in order to devise such a test. Critical values $c_{1-\alpha}$, $\alpha \in (0,1)$, can be obtained from the fact that

$$P\left(\sup_{t \in [0,1]} |B(t)| > x \right) = 1 - \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{(2k+1)^2 \pi^2}{8x^2} \right),$$

for $x > 0$, see e.g. [25]. We now reject the no–change null hypothesis in favor that a change has occured, if $V_n > c_{1-\alpha}$. In this case, the unknown (first) change-point, i.e. the onset, is estimated canonically by

$$\widehat{k}_n = \min\{k \le n : |\mathcal{D}_n(k/n)| \ge |\mathcal{D}_n(\ell/n)|,\ \ell = 1, \dots, n\}.$$

If $\mathbf{\Sigma}_n^{(0)}$ and α_n are unknown, they have to be estimated from a sufficiently large learning sample satisfying the no-change null hypothesis. Assuming that such a learning sample is given is, however, standard in the change-point literature, see e.g. [7] where it has been named *non-contamination assumption*. Alternative procedures that avoid the estimation of $\mathbf{\Sigma}_n^{(0)}$ will be discussed elsewhere in greater detail.

ACKNOWLEDGEMENTS

Parts of this paper have been written during a research stay of the first author as visiting professor at Université catholique de Louvain, Louvain-la-Neuve and a visit at the Department of Statistics of Seoul National University, Seoul, South Korea. He thanks Rainer von Sachs and Sangyeol Lee for their hospitality. Rainer von Sachs gratefully acknowledges funding by contract Projet d'Actions de Recherche Concertées" No. 12/17-045 of the ,,Communauté française de Belgique" and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy). He wants to thank the Institute of Stochastics of the RWTH Aachen (Prof. A. Steland and his group) for the warm hospitality during his sabbatical stay in autumn 2013, and part of this sabbatical has generously been funded by the Université catholique de Louvain.

REFERENCES

[1] Alexander Aue and Lajos Horváth. Structural breaks in time series. *J. Time Series Anal.*, 34(1):1–16, 2013.

[2] Peter J. Bickel and Elizaveta Levina. Covariance regularization by thresholding. *Ann. Statist.*, 36(6):2577–2604, 2008.

[3] Hilmar Böhm and Rainer von Sachs. Structural shrinkage of nonparametric spectral estimators for multivariate time series. *Electronic Journal of Statistics*, 2:696–721, 2008.

[4] Hilmar Böhm and Rainer von Sachs. Shrinkage estimation in the frequency domain of multivariate time series. *J. Multivariate Anal.*, 100(5):913–935, 2009.

[5] Joshua Brodie, Ingrid Daubechies, Christine De Mol, Domenico Giannone, and Ignace Loris. Sparse and stable Markowitz portfolios. *Proceedings the National Academy of Sciences of the United States of America*, 106(30):12267–12272, 2009.

[6] Julian Chan, Lajos Horváth, and Marie Hušková. Darling-Erdős limit results for change-point detection in panel data. *J. Statist. Plann. Inference*, 143(5):955–970, 2013.

[7] Chia-Shang James Chu, Maxwell Stinchcombe, and Halbert White. Monitoring structural change. *Econometrica*, 64(5):1045–1065, 1996.

[8] Miklós Csörgő and Lajos Horváth. *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1997. With a foreword by David Kendall.

[9] Persi Diaconis and David Freedman. Asymptotics of graphical projection pursuit. *Ann. Statist.*, 12(3):793–815, 1984.

[10] Mark Fiecas, Jürgen Franke, Rainer von Sachs, and Joseph Tadjuidje. Shrinkage estimation for multivariate hidden Markov mixture models. *Université catholique de Louvain, ISBA Discussion Paper 2012/16; submitted and in revision*, 2014.

[11] Mark Fiecas and Rainer von Sachs. Data-driven shrinkage of the spectral density matrix of a high-dimensional time series. *Université catholique de Louvain, ISBA Discussion Paper 2013/44; submitted and in revision*, 2013.

[12] Marie Hušková and Zdeněk Hlávka. Nonparametric sequential monitoring. *Sequential Anal.*, 31(3):278–296, 2012.

[13] Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, LVIII, 2003.

[14] Iain M. Johnstone and Arthur Yu Lu. On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.*, 104(486):682–693, 2009.

[15] L. Jolliffe, N. Trendafilov, and M. Uddin. A modified principal component technique based on the lasso. *Journal of Computational and Graphical Statistics*, 12:531–547, 2003.

[16] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32:111–131, 1975.

[17] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(1):33–58, 1976.

[18] Michael A. Kouritzin. Strong approximation for cross-covariances of linear variables with long-range dependence. *Stochastic Process. Appl.*, 60(2):343–353, 1995.

[19] Oliver Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621, 2003.

[20] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.*, 88(2):365–411, 2004.

[21] H. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.

[22] Walter Philipp. A note on the almost sure approximation of weakly dependent random variables. *Monatsh. Math.*, 102(3):227–236, 1986.

[23] Alessio Sancetta. Sample covariance shrinkage for high dimensional dependent data. *J. Multivariate Anal.*, 99(5):949–967, 2008.

[24] Haipeng Shen and Jianhua Z. Huang. Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivariate Anal.*, 99(6):1015–1034, 2008.

[25] Galen R. Shorack. *Probability for statisticians*. Springer Texts in Statistics. Springer-Verlag, New York, 2000.

[26] Charles Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, pages 197–206. University of California Press, Berkeley and Los Angeles, 1956.

[27] Ansgar Steland. Sequential control of time series by functionals of kernel-weighted empirical processes under local alternatives. *Metrika*, 60(3):229–249, 2004.

[28] Ansgar Steland. Optimal sequential kernel detection for dependent processes. *J. Statist. Plann. Inference*, 132(1-2):131–147, 2005.

[29] Ansgar Steland. A surveillance procedure for random walks based on local linear estimation. *J. Nonparametr. Stat.*, 22(3-4):345–361, 2010.

[30] Ansgar Steland. *Financial Statistics and Mathematical Finance*. Springer, 2012.

[31] Ansgar Steland and Ewaryst Rafajłowicz. Decoupling change-point detection based on characteristic functions: methodology, asymptotics, subsampling and application. *J. Statist. Plann. Inference*, 145:49–73, 2014.

[32] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996.

[33] Robert Tibshirani. Regression shrinkage and selection via the lasso: a retrospective. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(3):273–282, 2011.

[34] Daniela M. Witten and Robert Tibshirani. Testing significance of features by lassoed principal components. *Ann. Appl. Stat.*, 2(3):986–1012, 2008.

[35] Daniela M. Witten, Robert Tibshirani, and Trevor Hastie. A penalized decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10:515–534, 2009.

[36] Wei Biao Wu and Wanli Min. On linear processes with dependent innovations. *Stochastic Process. Appl.*, 115(6):939–958, 2005.