INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

DISCUSSION PAPER

2016/38

Asymptotics for High–Dimensional Covariance Matrices and Quadratic Forms with Applications to the Trace Functional and Shrinkage

Steland, A. and R. von Sachs

Preprint, Institute of Statistics, RWTH Aachen University

ASYMPTOTICS FOR HIGH–DIMENSIONAL COVARIANCE MATRICES AND QUADRATIC FORMS WITH APPLICATIONS TO THE TRACE FUNCTIONAL AND SHRINKAGE

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
SEPTEMBER 19, 2016

We establish large sample approximations for an arbitray number of bilinear forms of the sample variance-covariance matrix of a high-dimensional vector time series using ℓ_1-bounded weighting vectors. Estimation of the asymptotic covariance structure is also discussed. The results hold true without any constraint on the dimension, the number of forms and the sample size or their ratios. Concrete and potential applications are widespread and cover high-dimensional data science problems such as projections onto sparse principal components or more general spanning sets as frequently considered, e.g. in classification and dictionary learning. As two specific applications of our results, we study in greater detail the asymptotics of the trace functional and shrinkage estimation of the covariance matrices. In shrinkage estimation, it turns out that the asymptotics differs for weighting vectors bounded away from orthogonaliy and nearly orthogonal ones in the sense that their inner product converges to 0.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 60F17; secondary 62E20 .
KEYWORDS: Brownian motion, Linear process, Long memory, Strong approximation, Quadratic form, Trace.

1. INTRODUCTION

A large number of procedures studied to analyze high–dimensional vector time series of dimension d_n depending on the sample size n relies on projections, e.g. by projecting the observed random vector onto a spanning set of a lower dimensional subspace of dimension L_n. Examples include sparse principal component analysis, see e.g. [20], in order to reduce dimensionality of data, sparse portfolio replication and index tracking as studied by [4], or dictionary learning, see [1], where one aims at representing input data by a sparse linear combination of the elements of a dictionary, frequently obtained as the union of several bases and/or historical data.

When studying projections, it is natural to study the associated bilinear form $\mathbf{v}_n'\widehat{\mathbf{\Sigma}}_n\mathbf{w}_n$, $\mathbf{v}_n, \mathbf{w}_n \in \mathbb{R}^{d_n}$, representing the dependence structure in terms of the projections' covariances. Here and throughout the paper $\widehat{\mathbf{\Sigma}}_n$ is the (uncentered) sample variance–covariance matrix. In order to conduct inference, large sample distributional approximations are needed. For a vector time series model given by correlated linear processes, we established in [17] a strong approximation by a Brownian motion for a single quadratic form, provided the weighting vectors are uniformly bounded in the ℓ_1-norm. It turned out that the result does not require any condition on the ratio of the dimension and the sample size, contrary to many asymptotic results in highdimensional statistics and probability.

In the present article, we study the more general case of an *increasing* number of quadratic forms as arising when projecting onto a sequence of subspaces whose dimension converges to ∞. Noting that the analysis of autocovariances of a stationary linear time series appears as a special case of our approach, there are a few recent results related to our work: [23] established a central limit theorem for a finite number of autocovariances, whereas in [22] the case of long memory series has been studied. [8] has studied the asymptotic theory for detecting a change in mean of a vector time series with growing dimension.

To treat the case of an increasing number of bilinear forms, we consider two related but different frameworks: The first framework uses a sequence of Euclidean spaces $\mathbb{R}^{d_n}$ equipped with the usual Euclidean norm. The second framework embeds those spaces in the sequence space ℓ_2 equipped with the ℓ_2–norm. It is shown that, in both frameworks, an increasing number of, say L_n, quadratic forms can be approximated by Brownian motions without any constraints on L_n, d_n and n apart from $n \to \infty$. One of our main results asserts that, for the assumed time series models, one can define, on a new probability space, equivalent versions and a Gaussian process $\mathcal{G}_n$ taking values in $C([0,1], \mathbb{R}^{L_n})$, such that

$$\sup_{t\in[0,1]} \frac{1}{\sqrt{nL_n}} \left| \left(\mathbf{v}_n^{(j)\prime} (\widehat{\boldsymbol{\Sigma}}_{\lfloor nt \rfloor} - E\widehat{\boldsymbol{\Sigma}}_{\lfloor nt \rfloor}) \mathbf{w}_n^{(j)} \right)_{j=1}^{L_n} - \mathcal{G}_n(t) \right| = o_P(1),$$

as $n \to \infty$, almost surely (a.s.), without any constraints on L_n, d_n.

We believe that those results have many applications in diverse areas, as indicated above. In this paper we study in some detail two direct applications: The first application considers the trace operator, which equals the trace matrix norm $\| \bullet \|_{tr}$ when applied to covariance matrices. We show that the trace of the sample covariance matrix, appropriately centered, can be approximated by a Brownian motion, a.s. on a new probability space, which also establishes the convergence rate

$$\left| \|\widehat{\boldsymbol{\Sigma}}_n\|_{tr} - \|E\widehat{\boldsymbol{\Sigma}}_n\|_{tr} \right| = O_P(n^{-1/2} d_n).$$

The second application elaborated in this paper is shrinkage estimation of a covariance matrix as studied in depth for i.i.d. sequences of high–dimensional random vectors as well as dependent vector time series, see by [12], [13] and [16] amongst others. In order to regularize the sample variance-covariance matrix, the shrinkage estimator considers a convex combination with a well–defined target that usually corresponds to a simple regular model. We consider the identity target, i.e. a multiple of the identity matrix I_n of dimension d_n. To the best of our knowledge, large sample approximations for those estimators have not yet been studied. We show that, uniformly in the shrinkage weight for the convex combination, a bilinear form given by the shrinkage estimator can be approximated by a Gaussian process when it is centered at the shrunken true covariance matrix using the same shrinkage weight. By uniformity, the result also holds for the widely used estimator of the optimal shrinkage weight. For this estimated optimal weight the convergence rate under quite general conditions is known. It turns out that, when comparing the matrices in terms of a natural pseudodistance induced by bilinear forms, the convergence rate carries over from the optimal weight's estimator. We also compare the shrinkage estimator using the estimated optimal weight with an oracle estimator using the unknown optimal weight. Last, we study the case of nearly (i.e. asymptotically) orthogonal vectors. As a consequence of the Kabatjanskii-Levenstein bound, see [18], this property allows to place much more unit vectors on the unit sphere. It turns out that for nearly orthogonal vectors the nonparametric part dominates in large samples, contrary to the situation for vectors bounded away from orthogonality.

The high-dimensional time series model of the paper is as follows. At time n, $n \in \mathbb{N}$, we observe a $d = d_n$ dimensional mean zero vector time series

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

whose coordinates are causal linear processes,

$$Y_{ni}^{(\nu)} = \sum_{j=0}^{\infty} c_{nj}^{(\nu)} \epsilon_{i-j}, \qquad i \in \mathbb{N}_0, \nu = 1, \dots, d_n, \tag{1.1}$$

where $\{\epsilon_t\}$ are independent mean zero error terms with $E(\epsilon_k^3) = 0$ for all k, possibly not identically distributed, with finite moments of the order $4+\delta$ for some $\delta > 0$. The coefficients $c_{nj}^{(\nu)}$ may depend on n and are therefore also allowed to depend on the dimension d_n. We impose the following growth condition.

Assumption (A): The coefficients $c_{nj}^{(\nu)}$ of the linear processes (1.1) satisfy

$$\sup_{n \in \mathbb{N}} \max_{1 \le \nu \le d_n} |c_{nj}^{(\nu)}|^2 \ll j^{-3/2-\theta} \tag{1.2}$$

for some $0 < \theta < 1/2$.

It is well known that Assumption (A) covers common classes of weakly dependent time series such as ARMA(p,q)-models as well as a wide range of long memory processes. We refer to [17] for a discussion.

Define the (centered) bilinear form

$$Q_n(\mathbf{v}_n, \mathbf{w}_n) = \sqrt{n}\mathbf{v}_n'(\widehat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n)\mathbf{w}_n, \qquad \mathbf{v}_n, \mathbf{w}_n \in \mathbb{R}^{d_n}, \tag{1.3}$$

where

$$\widehat{\mathbf{\Sigma}}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_{ni} \mathbf{Y}_{ni}' \qquad \text{and} \qquad \mathbf{\Sigma}_n = E\widehat{\mathbf{\Sigma}}_n. \tag{1.4}$$

The class of proper (sequences of) weighting vectors,

$$\mathbf{w}_n = (w_{n1}, \dots, w_{nd_n})', \qquad n \ge 1,$$

studied throughout the paper is the set $\mathcal{W}$ of those sequences $\{\mathbf{w}_n : n \ge 1\}$, $\mathbf{w}_n \in \mathbb{R}^{d_n}$, $n \ge 1$, which have uniformly bounded ℓ_1-norm in the sense that

$$\sup_{n \in \mathbb{N}} \|\mathbf{w}_n\|_{\ell_1} = \sup_{n \in \mathbb{N}} \sum_{\nu=1}^{d_n} |w_{n\nu}| < \infty. \tag{1.5}$$

ℓ_1-weighting vectors naturally arise in various applications such as sparse principal component analysis as, see e.g. [20], or sparse financial portfolio selection as studied by [4]. For a more detailed discussion we refer to [17].

The rest of the paper is organized as follows. In Section 2, we introduce the partial sums and partial sum processes associated to an increasing number of bilinear forms and establish the strong and weak approximation theorems for those bilinear forms. The application to the trace functional is discussed in Section 3. The large sample approximations for shrinkage estimators of covariance matrices are studied in depth in Section 4.

2. LARGE SAMPLE APPROXIMATIONS FOR HIGH–DIMENSIONAL BILINEAR FORMS

2.1. *Definitions and review*

Let us define the matrix–valued partial sums

$$\widehat{\boldsymbol{\Sigma}}_{nk} = \left(\sum_{i=1}^{k} Y_i^{(\nu)} Y_i^{(\mu)} \right)_{1 \le \nu,\mu \le d_n}, \tag{2.1}$$

$$\boldsymbol{\Sigma}_{nk} = \left(\sum_{i=1}^{k} E Y_i^{(\nu)} Y_i^{(\mu)} \right)_{1 \le \nu,\mu \le d_n}, \tag{2.2}$$

for $n, k \geq 1$, and put

$$D_{nk}(\mathbf{v}_n, \mathbf{w}_n) = \mathbf{v}_n' (\widehat{\boldsymbol{\Sigma}}_{nk} - \boldsymbol{\Sigma}_{nk}) \mathbf{w}_n, \qquad n, k \geq 1, \tag{2.3}$$

for two sequences of weighting vectors $\{\mathbf{v}_n\}, \{\mathbf{w}_n\} \in \mathcal{W}$. The associated càdlàg processes will be denoted by

$$\mathcal{D}_n(t; \mathbf{v}_n, \mathbf{w}_n) = n^{-1/2} D_{n,\lfloor nt \rfloor}(\mathbf{v}_n, \mathbf{w}_n) = \mathbf{v}_n' n^{-1/2} (\widehat{\boldsymbol{\Sigma}}_{n,\lfloor nt \rfloor} - \boldsymbol{\Sigma}_{n,\lfloor nt \rfloor}) \mathbf{w}_n, \qquad t \in [0,1], n \geq 1. \tag{2.4}$$

Especially, we have

$$\mathcal{D}_n(1) = \mathcal{D}_n(1; \mathbf{v}_n, \mathbf{w}_n) = \mathbf{v}_n' \sqrt{n} (\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n) \mathbf{w}_n, n \geq 1, \tag{2.5}$$

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

For some sequence of standard Brownian motions, $\{B_n(t) : t \in [0, \infty)\}$, $n \geq 1$, and any constant $N > 0$ we can introduce the rescaled version $\{N^{-1/2} B_N(tN) : s \in [0,1]\}$ called the *[0,1]–version of* B_n. In [17] the following result on the asymptotics of a single bilinear form for a uniformly bounded ℓ_1-projections is shown:

THEOREM 2.1 *Suppose $\mathbf{Y}_{ni}$, $1 \leq i \leq n$, $n \geq 1$, is a vector time series according to model (1.1) that satisfies Assumption (A). Let $\mathbf{v}_n$ and $\mathbf{w}_n$ be weighting vectors with uniformly bounded ℓ_1–norm in the sense of (1.5). Then, for each $n \in \mathbb{N}$, there exists equivalent versions of $D_{nk}(\mathbf{v}_n, \mathbf{w}_n)$ and $D_n(t; \mathbf{v}_n, \mathbf{w}_n)$, $t \geq 0$, again denoted by $D_{nk}(\mathbf{v}_n, \mathbf{w}_n)$ and $D_n(t; \mathbf{v}_n, \mathbf{w}_n)$, and a standard Brownian motion $\{W_n(t) : t \geq 0\}$, which depends on $(\mathbf{v}_n, \mathbf{w}_n)$, i.e. $W_n(t) = W_n(t; \mathbf{v}_n, \mathbf{w}_n)$, both defined on some probability space $(\Omega_n, \mathcal{F}_n, P_n)$, such that for some $\lambda > 0$ and a constant C_n*

$$|D_{nt}(\mathbf{v}_n, \mathbf{w}_n) - \alpha_n(\mathbf{v}_n, \mathbf{w}_n) W_n(t)| \leq C_n t^{1/2-\lambda}, \qquad \textit{for all } t > 0 \textit{ a.s.,} \tag{2.6}$$

where $\alpha_n(\mathbf{v}_n, \mathbf{w}_n)$ is defined in (A.7). If $C_n n^{-\lambda} = o(1)$, as $n, t \to \infty$, this implies the strong approximation

$$\sup_{t\in[0,1]} |\mathcal{D}_n(t; \mathbf{v}_n, \mathbf{w}_n) - \alpha_n(\mathbf{v}_n, \mathbf{w}_n) W_n(\lfloor nt \rfloor / n)| = o(1), \qquad a.s., \tag{2.7}$$

as $n \to \infty$, for the $[0,1]$–version of W_n as well as the CLT

$$|\mathcal{D}_n(1; \mathbf{v}_n, \mathbf{w}_n) - \alpha_n(\mathbf{v}_n, \mathbf{w}_n) W_n(1)| = o(1), \qquad a.s., \tag{2.8}$$

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

A multivariate version for $L \in \mathbb{N}$ bilinear forms, which approximates

$$\left(\mathcal{D}_n(\bullet; \mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)})\right)_{j=1}^{L}$$

by a L–dimensional Brownian motion, has been shown in [17, Th. 4.2]. This result allows to consider the dependence structure which arises when mapping $\mathbf{Y}_n = \mathbf{Y}_{nn}$ onto the subspace $\mathcal{P}_L = \operatorname{span}\{\mathbf{w}_n^{(j)} : j = 1, \dots, L\}$ spanned by L weighting vectors,

$$\mathbf{w}_n^{(j)} \in \mathcal{W}, \qquad j = 1, \dots, L.$$

We have the canonical mapping, called *projection (onto $\mathcal{P}_L$)* in the sequel,

$$\mathbf{Y}_n \mapsto \mathbf{P}_n \mathbf{Y}_n, \qquad \mathbf{P}_n = (\mathbf{w}_n^{(1)}, \dots, \mathbf{w}_n^{(L)})', \tag{2.9}$$

which represents the orthogonal projection onto $\mathcal{P}_L$, if $\mathbf{w}_n^{(1)}, \dots, \mathbf{w}_n^{(L)}$ are orthonormal. The associated variance-covariance matrix is

$$\operatorname{Cov}(\mathbf{P}_n \mathbf{Y}_n) = \left(\operatorname{Cov}(\mathbf{w}_n^{(j)} \mathbf{Y}_n, \mathbf{w}_n^{(k)} \mathbf{Y}_n)\right)_{1\le j,k\le L} = \left(\mathbf{w}_n^{(j)\prime} \mathbf{\Sigma}_n \mathbf{w}_n^{(k)}\right)_{1\le j,k\le L}.$$

If the $\mathbf{w}_n^{(j)}$'s are eigenvectors of $\mathbf{\Sigma}_n$, then $\operatorname{Cov}(\mathbf{P}_n \mathbf{Y}_n)$ is a diagonal matrix. But that property is lost for more general spanning vectors. Given the sample $\mathbf{Y}_{n1}, \dots, \mathbf{Y}_{nn}$ of d_n-dimensional random vectors, the canonical nonparametric statistical estimators of $\mathbf{\Sigma}_n$ and $\operatorname{Cov}(\mathbf{P}_n \mathbf{Y}_n)$ are $\widehat{\mathbf{\Sigma}}_n$ as defined in (1.4) and

$$\widehat{\operatorname{Cov}}(\mathbf{P}_n \mathbf{Y}_n) = \left(\mathbf{w}_n^{(j)\prime} \widehat{\mathbf{\Sigma}}_n \mathbf{w}_n^{(k)}\right)_{1\le j,k\le L}.$$

The entries of $\widehat{\operatorname{Cov}}(\mathbf{P}_n \mathbf{Y}_n)$ consist of bilinear forms as studied in Theorem 2.1, and for fixed L its multivariate extension suffices to study the dependence structure of the projection onto $\mathcal{P}_L$. This no longer holds, if L is allowed to grow as the sample size increases, i.e. when studying the case $L = L_n \to \infty$, as $n \to \infty$. Indeed, the treatment of that high–dimensional situation is much more involved. As we shall see, it requires a different scaling and a more involved mathematical framework: The strong approximations we establish in this paper take place in the Euclidean space $\mathbb{R}^{L_n}$ of growing dimension and the infinite-dimensional Hilbert space ℓ_2, respectively.

Thus, to go beyond the case of a finite number of bilinear forms we now consider

$$Q_n(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}) = \mathbf{v}_n^{(j)\prime} \widehat{\mathbf{\Sigma}}_n \mathbf{w}_n^{(j)}, \qquad j = 1, \dots, L_n,$$

where $(\{\mathbf{v}_\ell^{(j)}\}_{\ell=1}^\infty, \{\mathbf{w}_\ell^{(j)}\}_{\ell=1}^\infty) \in \mathcal{W} \times \mathcal{W}$, $j = 1, \dots, L_n$, are L_n pairs of uniformly ℓ_1-bounded sequences of weighting vectors and L_n may tend to infinity as $n \to \infty$. We are interested in the joint asymptotics of the centered and scaled versions of the corresponding statistics (2.5) given by

$$\mathcal{D}_{nj} = L_n^{-1/2} \mathcal{D}_n(1; \mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}), \qquad j = 1, \dots, L_n, \tag{2.10}$$

and the associated sequential càdlàg processes

$$\mathcal{D}_{nj}(t) = L_n^{-1/2} \mathcal{D}_n(t; \mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}), \qquad t \in [0,1], \qquad j = 1, \dots, L_n, \tag{2.11}$$

cf. (2.4). The additional factor $L_n^{-1/2}$ anticipates the right scaling to obtain a large sample approximation.

Further, we are interested in studying weighted averages where averaging takes place over all L_n forms and all sample sizes $n \in \mathbb{N}$. Let λ_n be the weight for sample size n and μ_ρ the weight for the ρth quadratic form associated to a pair of sequences of weighting vectors $\{\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)} : n \geq 1\}$ for $\rho = 1, \dots, L_n$, $n \geq 1$. Define for $k \geq 1$

$$D_k(\{(\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)})\}) = \sum_{n,m=1}^\infty \lambda_n \lambda_m \sum_{\rho=1}^{L_n} \sum_{\sigma=1}^{L_m} \mu_\rho \mu_\sigma \mathbf{v}_n^{(\rho)\prime} (\widehat{\mathbf{\Sigma}}_{nmk} - \mathbf{\Sigma}_{nmk}) \mathbf{w}_m^{(\sigma)}, \tag{2.12}$$

where

$$\widehat{\mathbf{\Sigma}}_{nmk} = \sum_{i \leq k} \mathbf{Y}_{ni} \mathbf{Y}_{mi}', \tag{2.13}$$

$$\mathbf{\Sigma}_{nmk} = E(\widehat{\mathbf{\Sigma}}_{nmk}), \tag{2.14}$$

for $n, m \geq 1$. Notice the relations

$$\widehat{\mathbf{\Sigma}}_{nnk} = \widehat{\mathbf{\Sigma}}_{nk}, \qquad \widehat{\mathbf{\Sigma}}_{nn} = n \widehat{\mathbf{\Sigma}}_n$$

between (1.4), (2.1) and (2.13). $D_k(\{(\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)})\})$ depends on all weights $\{v_{n\nu}^{(\rho)}, w_{n\nu}^{(\rho)}, \lambda_n, \mu_\rho : 1 \leq \nu \leq d_n, 1 \leq \rho \leq L_n, n \geq 1\}$ but is measurable w.r.t $\mathcal{G}_k = \sigma(\mathbf{Y}_{ni} : n \in \mathbb{N}, i \leq k)$. Now, for any sample size M we may consider the associated càdlàg process associated to (2.12)

$$\mathcal{D}_M(t; \{(\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)})\}) = M^{-1/2} D_{\lfloor Mt \rfloor}(\{(\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)})\}), \qquad t \in [0,1]. \tag{2.15}$$

2.2. *Preliminaries*

Before proceeding, recall the following facts on the Hilbert space ℓ_2 and strong approximations in Hilbert spaces. The space ℓ_2 of sequences $f = (f_j)_j$ with $\sum_j f_j^2 < \infty$ is a separable Hilbert space when equipped with the inner product $(f, g) = \sum_j f_j g_j$, $f = (f_j)_j, g = (g_j)_j \in \ell_2$, and the induced norm $\|f\|_\ell = \sqrt{(f, f)}$.

Sufficient conditions for a strong approximation of partial sums of dependent random elements taking values in a separable Hilbert space require the control of the associated conditional covariance operator. Denote the underlying probability space by $(\Omega, \mathcal{F}, P)$. Let

$X = (X_j)_j$ be a random element defined on $(\Omega, \mathcal{F}, P)$ taking values in ℓ_2 with $E\|X\|_{\ell_2}^2 = \sum_j E(X_j^2) < \infty$. The covariance operator $C_X : \ell_2 \to \ell_2$ associated to X is defined by

$$C_X(f) = E[(f,X)X] = \left(\sum_j f_j E(X_j X_k) \right)_k, \qquad f = (f_j)_j \in \ell_2.$$

To any σ-field $\mathcal{A}$ we may associate the conditional covariance operator of X given $\mathcal{A}$,

$$C_X(f|\mathcal{A}) = E[(f,X)X|\mathcal{A}] = \left(\sum_j f_j E(X_j X_k|\mathcal{A}) \right)_k, \qquad f = (f_j)_j \in \ell_2. \tag{2.16}$$

Covariance operators are symmetric positive linear operators with operator norm

$$\|C_X(\bullet|\mathcal{A})\| = \sup_{f \in \ell_2 : \|f\|_{\ell_2}=1} \|(C_X(f), f)\|_{\ell_2}. \tag{2.17}$$

For further properties and discussion see [3].

A strong invariance principle in ℓ_2 deals with the a.s. approximation of partial sums of ℓ_2–valued random elements by a Brownian motion in ℓ_2. Recall that a random element $B = \{B(t) : t \in [0,1]\}$ with values in $C([0,1], \ell_2)$ is called Brownian motion in ℓ_2, if

(i) $B(0) = 0 \in \ell_2$,

(ii) for all $0 \le t_1 < \cdots < t_n \le 1$ the increments $B(t_{i+1}) - B(t_i)$, $i = 1, \ldots, n-1$, are independent and

(iii) for all $0 \le s, t \le 1$ the increment $B(t) - B(s)$ is Gaussian with mean 0 and covariance operator $\min(s,t)K$ for some nonnegative linear and self-adjoint operator K on ℓ_2 such that $\sum_{i=1}^\infty (Ke_i, e_i) < \infty$, where $\{e_i\}$ is some orthonormal system for ℓ_2.

If $K = C_X$ for some random element X, B is the Brownian motion generated by X. The definition for a general separable Hilbert space is analogous.

A strong invariance principle or strong approximation for a sequence $\zeta_1, \zeta_2, \ldots$ of random elements taking values in an arbitrary separable Hilbert space H with inner product $(\bullet, \bullet)$ and induced norm $|\bullet|$ asserts that they can be redefined on a rich enough probability space such that there exists a Brownian motion B with values in H and covariance operator C_ζ such that, a.s.,

$$\left| \sum_{j \le t} \zeta_j - B(t) \right| \le c t^{1/2 - \lambda}, \tag{2.18}$$

for constants $\lambda > 0$ and $c < \infty$, if the dimension of H is finite, and,

$$\left| \sum_{j \le t} \zeta_j - B(t) \right| = o(\sqrt{t \log(\log t)}), \tag{2.19}$$

as $t \to \infty$, a.s., if H is infinite dimensional.

Throughout the paper we write, for two arrays $\{a_{n',m'}\}$ and $\{b_{n',m'}\}$ of real numbers, $a_{n',m'} \overset{n',m'}{\ll} b_{n',m'}$, if there exists a constant c such that $a_{n',m'} \le c b_{n',m'}$ for all n', m'.

2.3. *Large sample approximations*

We aim at showing a strong approximation for the $D([0,1];\ell_2)$-valued processes

$$\mathcal{D}_n = (\mathcal{D}_{nj})_{j=1}^{L_n}, \qquad n \geq 1,$$

where the coordinate processes $\mathcal{D}_{nj}$ are given by

$$\mathcal{D}_{nj}(t) = \frac{1}{\sqrt{nL_n}} D_{n,\lfloor nt \rfloor}(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}), \qquad t \in [0,1], \quad n \geq 1,$$

with $D_{nk}(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}) = \mathbf{v}_n^{(j)\prime}(\widehat{\mathbf{\Sigma}}_{nk} - \mathbf{\Sigma}_{nk})\mathbf{w}_n^{(j)}$ for $j = 1, \dots, L_n$, $n \geq 1$, cf. (2.3) and (2.10). The above processes can be expressed as partial sums.

Lemma 2.1 *We have the representation*

$$D_{nk} = \sum_{i=1}^{k} \xi_i^{(n)}, \tag{2.20}$$

for $k = 1, \dots, n$, $n \geq 1$, leading to

$$\mathcal{D}_n(t) = \frac{1}{\sqrt{nL_n}} \sum_{i=1}^{\lfloor nt \rfloor} \xi_i^{(n)}, \qquad t \in [0,1], n \geq 1, \tag{2.21}$$

where the random elements $\xi_i^{(n)}$ are defined in (2.37).

To introduce the conditional covariance operators associated to $\mathcal{D}_n$, denote the filtration $\mathcal{F}_m = \sigma(\epsilon_i : i \leq m)$, $m \in \mathbb{Z}$, and define

$$\mathcal{C}^{(n)}(f|\mathcal{F}_0) = E[(f, \mathcal{D}_n(1))\mathcal{D}_n(1) \,|\, \mathcal{F}_0], \qquad f \in \ell_2, n \geq 1.$$

Let us also introduce the unconditional covariance operator

$$\mathcal{C}^{(n)}(f) = E[(f, Z_n)Z_n], \qquad f \in \ell_2,$$

where $Z_n = (Z_{nj})_{j=1}^{L_n}$, $n \geq 1$, with random variables Z_{nj}, $1 \leq j \leq L_n$, satisfying $E(Z_{nj}) = 0$ and

$$E(Z_{nj}Z_{nk}) = L_n^{-1}\beta_n^2(j,k),$$

for $j, k = 1, \dots, L_n$. Here

$$\beta_n^2(j,k) = \beta_n^2(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}, \mathbf{v}_n^{(k)}, \mathbf{w}_n^{(k)}) \tag{2.22}$$

are the quantities introduced in (A.9), the asymptotic covariance parameters of the bilinear forms corresponding to the pairs $(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)})$ and $(\mathbf{v}_n^{(k)}, \mathbf{w}_n^{(k)})$, $1 \leq j, k \leq L_n$.

The following technical but crucial result establishes the convergence of $\mathcal{C}^{(n)}(\bullet|\mathcal{F}_0) - \mathcal{C}^{(n)}(\bullet)$ in the operator semi–norm in expectation and provides us with a convergence rate.

THEOREM 2.2 *Suppose* $(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)})$, $j = 1, \dots, L_n$, *have uniformly bounded* ℓ_1*-norms,*

$$\sup_{n \geq 1} \max_{1 \leq j \leq L_n} \max\{\|\mathbf{v}_n^{(j)}\|_{\ell_1}, \|\mathbf{w}_n^{(j)}\|_{\ell_1}\} \leq C < \infty,$$

for some constant C. *Let*

$$S_{n',m'}^{(n)} = \frac{1}{\sqrt{n' L_n}} \sum_{k=m'+1}^{m'+n'} \xi_k^{(n)}, \qquad m', n' \geq 1, n \geq 1, \tag{2.23}$$

with $\xi_k^{(n)}$ *defined by (2.37). Define*

$$\mathcal{C}_{n',m'}^{(n)}(f|\mathcal{F}_{m'}) = E[(f, S_{n',m'}^{(n)}) S_{n',m'}^{(n)} | \mathcal{F}_{m'}],$$

for $f \in \ell_2$. *Then*

$$E\|\mathcal{C}_{n',m'}^{(n)}(\bullet|\mathcal{F}_m) - \mathcal{C}^{(n)}(\bullet)\| \overset{n',m'}{\ll} (n')^{-\theta/2},$$

where $\| \bullet \|$ *denotes the operator norm defined in (2.17).*

We are now in a position to formulate the first main result on the large sample approximations of L_n bilinear forms when L_n converges to infinity, in terms of the ℓ_1– as well as the ℓ_2–norm. The results holds true under the weak assumption that the weighting vectors have uniformly bounded ℓ_1 norm.

THEOREM 2.3 *Let* $\mathbf{Y}_{ni}$, $1 \leq i \leq n$, *be a vector time series following model (1.1) and satisfying Assumption (A). Suppose that* $(\{\mathbf{v}_\ell^{(1)}\}_{\ell=1}^\infty, \{\mathbf{w}_\ell^{(1)}\}_{\ell=1}^\infty), \dots, (\{\mathbf{v}_n^{(L_n)}\}_{\ell=1}^\infty, \{\mathbf{w}_n^{(L_n)}\}_{\ell=1}^\infty) \in \mathcal{W} \times \mathcal{W}$, $n \geq 1$, *have uniformly bounded* ℓ_1*-norm, i.e.*

$$\sup_{n \geq 1} \max_{1 \leq j \leq L_n} \max\{\|\mathbf{v}_n^{(j)}\|_{\ell_1}, \|\mathbf{w}_n^{(j)}\|_{\ell_1}\} \leq C,$$

for some constant $C < \infty$. *Then all processes can be redefined on a rich enough probability space, such that there exists, for each* n, *a Brownian motion of dimension* L_n,

$$B_n(t) = B_n(t; \{(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}) : 1 \leq j \leq L_n\}), \qquad t \geq 0,$$

with coordinates $B_n(t)_j$, $j = 1, \dots, L_n$, *and covariance function given by*

$$E(B_n(s)_j B_n(t)_k) = \min(s,t) L_n^{-1} \beta_n^2(j,k), \tag{2.24}$$

for $1 \leq j, k \leq L_n$ *and* $s, t \geq 0$, *such that the following assertions hold true.*

(i) In the Euclidean space $\mathbb{R}^{L_n}$ *we have the strong approximation*

$$\|L_n^{-1/2} D_{nt} - B_n(t)\|_{\mathbb{R}^{L_n}} = \left\| \sum_{i=1}^{t} L_n^{-1/2} \xi_i^{(n)} - B_n(t) \right\|_{\mathbb{R}^{L_n}} \leq C_n t^{-1/2-\lambda_n}, \tag{2.25}$$

a.s, for constants $C_n < \infty$ *and* $\lambda_n > 0$, *where* λ_n *depends only on* L_n, δ *and* θ.

Provided

$$C_n n^{-\lambda_n} = o(1), \tag{2.26}$$

as $n \to \infty$, the following assertions hold.

(ii) With respect to the ℓ_2–norm we have

$$\sup_{t\in[0,1]} \sum_{j=1}^{L_n} \left| \mathcal{D}_{nj}(t) - \overline{B}_n\left(\frac{\lfloor nt \rfloor}{n}\right)_j \right|^2 = o(1)$$

as $n \to \infty$, for the $[0,1]$–version $\overline{B}_n$ of B_n.

(iii) With respect to the ℓ_1–norm we have

$$\sup_{t\in[0,1]} \frac{1}{\sqrt{L_n}} \sum_{j=1}^{L_n} \left| \mathcal{D}_{nj}(t) - \overline{B}_n\left(\frac{\lfloor nt \rfloor}{n}\right)_j \right| = o(1),$$

as $n \to \infty$, for the $[0,1]$–version $\overline{B}_n$ of B_n, and with respect to the maximum norm

$$\sup_{t\in[0,1]} \max_{j \le L_n} \left| \mathcal{D}_{nj}(t) - \overline{B}_n\left(\frac{\lfloor nt \rfloor}{n}\right)_j \right| = o(1),$$

as $n \to \infty$.

(iv) Let $\{\lambda_n : n \in \mathbb{N}\}$ and $\{\mu_\rho : \rho \in \mathbb{N}\}$ be ℓ_1–weights. Then there exist constants $\lambda > 0$ and $C < \infty$ and $\alpha(\{\lambda_n\}_{n=1}^\infty, \{\mu_\rho\}_{\rho=1}^\infty, \{(\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)})\}_{n,\rho=1}^\infty) \ge 0$, such that for equivalent versions and a standard Brownian motion B on $[0,\infty)$, defined on a new probability space,

$$|D_t(\{(\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)})\}) - \alpha(\{\lambda_n\}_{n=1}^\infty, \{\mu_\rho\}_{\rho=1}^\infty, \{(\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)})\}_{n,\rho=1}^\infty) B(t)| \le C t^{1/2-\lambda}, \tag{2.27}$$

a.s., for all $t > 0$. Further, for any sample size M

$$\sup_{t\in[0,1]} |\mathcal{D}_M(t; \{(\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)})\}) - \alpha(\{\lambda_n\}_{n=1}^\infty, \{\mu_\rho\}_{\rho=1}^\infty, \{(\mathbf{v}_n^{(\rho)}, \mathbf{w}_n^{(\rho)})\}_{n,\rho=1}^\infty) \overline{B}_M(t)| \le C M^{-\lambda}, \tag{2.28}$$

a.s., for the $[0,1]$–version $\overline{B}_M$ of B.

Remark 2.1 *The Brownian motions can be constructed such that*

$$\sup_{t\in[0,1]} \left| \overline{B}_n\left(\frac{\lfloor nt \rfloor}{n}\right)_j - L_n^{-1/2} \alpha_n(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}) W_n\left(\frac{\lfloor nt \rfloor}{n}; \mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}\right) \right| = o(1), \tag{2.29}$$

as $n \to \infty$, a.s., if (2.26) holds, where $W_n(\bullet; \mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)})$ is as in Theorem 2.1, for $j = 1, \dots, L_n$.

Due to assertion (iv) of the above theorem we may conjecture that (2.26) holds, cf. the discussion in [17], but we have neither a proof nor a counterexample. The following result studies the relevant processes in the infinite–dimensional space ℓ_2 and yields an approximation in probability taking into account the additional factor $\log\log(n)$.

THEOREM 2.4 *Suppose that the assumptions of Theorem 2.3 hold. In the Hilbert space ℓ_2 we have the strong approximation*

$$\|L_n^{-1/2} D_{nt} - B_n(t)\|_{\ell_2} = \left\| \sum_{k=1}^{t} L_n^{-1/2} \xi_k^{(n)} - B_n(t) \right\|_{\ell_2} = o(\sqrt{t \log(\log t)}), \tag{2.30}$$

as $t \to \infty$, a.s. There exists a sequence $\{\delta_n : n \geq 1\} \subset \mathbb{N}$ such that with $N = \lceil n \log\log(n) \rceil$

$$\max_{\delta_n \leq k \leq n} \left\| \frac{1}{\sqrt{N L_n}} \sum_{i=1}^{k} \xi_i^{(n)} - \overline{B}_N\left(\frac{k}{N}\right) \right\|_{\ell_2} = o_P(1), \tag{2.31}$$

$n \to \infty$, for the $[0,1]$–version $\overline{B}_N(t) = N^{-1/2} B_n(tN)$. In other words,

$$\max_{\delta_n \leq k \leq n} \left\| \sqrt{\frac{n}{N}} \mathcal{D}_n\left(\frac{k}{n}\right) - \overline{B}_N\left(\frac{k}{N}\right) \right\|_{\ell_2} = o_P(1), \tag{2.32}$$

or equivalently

$$\sup_{\frac{\delta_n}{n} \leq t \leq 1} \left\| \sqrt{\frac{n}{N}} \mathcal{D}_n(t) - \overline{B}_N\left(\frac{\lfloor nt \rfloor}{N}\right) \right\|_{\ell_2} = o_P(1), \tag{2.33}$$

as $n \to \infty$.

The above result eliminates the condition (2.26), but we have no detailed information about the sequence δ_n.

2.4. *Proofs*

PROOF OF LEMMA 2.1: We argue as in the proof of Theorem 2.1 given in [17], where it was shown that the partial sum (2.3) associated to a single bilinear form $Q(\mathbf{v}_n, \mathbf{w}_n)$ attains the representation

$$D_{nk}(\mathbf{v}_n, \mathbf{w}_n) = \sum_{i \leq k} \xi_i^{(n)}(\mathbf{v}_n, \mathbf{w}_n) \tag{2.34}$$

with Gaussian random variables

$$\xi_i^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = Y_{ni}(\mathbf{v}_n) Y_{ni}(\mathbf{w}_n) - E(Y_{ni}(\mathbf{v}_n) Y_{ni}(\mathbf{w}_n)), \qquad i = 1, 2, \ldots, n, \ n \geq 1, \tag{2.35}$$

for linear processes

$$Y_{ni}(\mathbf{v}_n) = \sum_{j=0}^{\infty} c_{nj}^{(v)} \epsilon_{i-j}, \qquad Y_{ni}(\mathbf{w}_n) = \sum_{j=0}^{\infty} c_{nj}^{(w)} \epsilon_{i-j}, \qquad i \in \mathbb{Z}, n \in \mathbb{N},$$

with coefficients

$$c_{nj}^{(v)} = \sum_{\nu=1}^{d_n} v_\nu c_{nj}^{(\nu)}, \qquad c_{nj}^{(w)} = \sum_{\nu=1}^{d_n} w_\nu c_{nj}^{(\nu)}, \tag{2.36}$$

for $j \geq 0$ and $n \geq 1$. For L_n pairs of weighting vectors $(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)})$, $j = 1, \dots, L_n$, we consider the corresponding partial sum process where the summands are the L_n-dimensional vectors

$$\xi_i^{(n)} = \left(\xi_i^{(n)}(j) \right)_{j=1}^{L_n}, \qquad \xi_i^{(n)}(j) = \xi_i^{(n)}(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}),\ j = 1, \dots, L_n, \tag{2.37}$$

for $i = 1, 2, \dots,$ which we, however, also interpret as random elements taking values in ℓ_2. This completes the proof. *Q.E.D.*

[15, Th. 1] asserts that (2.18) and (2.19), respectively, hold, if the following conditions for the scaled partial sums $S_{n'}(m') = (n')^{-1/2} \sum_{k=m'+1}^{m'+n'} \zeta_k$, $k, m', n' \geq 0$, are satisfied:

(I) $\sup_{j \geq 1} E|\zeta_j|^{2+\delta} < \infty$ for some $\delta > 0$.

(II) For some $\varepsilon > 0$

$$E|E(S_{n'}(m'))|\mathcal{F}_{m'})| \overset{n',m'}{\ll} (n')^{-\varepsilon}.$$

(III) There exists a covariance operator C such that the conditional covariance operators

$$C_{n'}(f|\mathcal{F}_{m'}) = E[(f, S_{n'}(m'))S_{n'}(m')|\mathcal{F}_{m'}), \qquad f \in H,$$

converge in the operator semi-norm $\| \bullet \|$ to $C(f)$ in expectation with rate $(n')^{-\theta}$, i.e.

$$E\|C_{n'}(\bullet|\mathcal{F}_{m'}) - C(\bullet)\| \overset{n',m'}{\ll} (n')^{-\theta},$$

for some $\theta > 0$.

As shown by [5], the strong invariance principle (2.19) also holds true for strictly stationary sequences taking values in a separable Hilbert space, which possess a finite moment of order $2+\delta'$, $\delta' > 0$, and are strong mixing with mixing coefficients satisfying $\alpha(k) = O(k^{-(1+\varepsilon)(1+2/\delta')})$, for some $\varepsilon > 0$. The above conditions are, however, more convenient when studying linear processes. [21] has studied strong invariance principles for a univariate nonlinear time series using the physical dependence measure, which is easy to verify for linear processes. Extensions to vector-valued time series (of fixed dimension) have been provided by [14]. We rely on the conditions of [15], since they allow to study time series of growing dimension and taking values in the infinite-dimensional space ℓ_2 in a relatively straightforward way.

As a preparation for the proof of Theorem 2.2, we need the following lemma dealing with the uniform convergence of unconditional and conditional covariances of the approximating martingales defined by

$$M_{m'}^{(n)}(\nu) = \sum_{k=0}^{m'} \left\{ \widetilde{f}_{0,0}^{(n)}(\nu)(\epsilon_k^2 - \sigma_k^2) + \epsilon_k \sum_{l=1}^{\infty} \widetilde{f}_{l,0}^{(n)}(\nu)\epsilon_{k-l} \right\}, \qquad m' \geq 0, \tag{2.38}$$

where for brevity $\widetilde{f}_{l,i}^{(n)}(\nu) = \widetilde{f}_{l,i}^{(n)}(\mathbf{v}_n^{(\nu)}, \mathbf{w}_n^{(\nu)})$, $l, i = 0, 1, 2, \dots,$ $\nu = 1, \dots, d_n$.

LEMMA 2.2 *Under Assumption (A) and if $E(\epsilon_k^3) = 0$ for all k*

(2.39)

$$\sup_{n \geq 1} \left\| \sup_{1 \leq \nu,\mu} \left| E[(M_{m'+n'}^{(n)}(\nu) - M_{m'}^{(n)}(\nu))(M_{m'+n'}^{(n)}(\mu) - M_{m'}^{(n)}(\mu))|\mathcal{F}_{m'}] - n'\beta_n(\nu,\mu) \right| \right\|_{L_1} \overset{n',m'}{<<} (n')^{1-\theta},$$

which implies

$$\sup_{n\geq 1}\sup_{n',m'\geq 1}(n')^{-(1-\theta)}\sup_{1\leq\nu}E[(M^{(n)}_{m'+n'}(\nu)-M^{(n)}_{m'}(\nu))^2|\mathcal{F}_{m'}]<\infty, \tag{2.40}$$

a.s. Further,

(2.41)

$$(n')^{-1}\sup_{n\geq 1}\sup_{1\leq\nu,\mu}\left|E(M^{(n)}_{m'+n'}(\nu)-M^{(n)}_{m'}(\nu))(M^{(n)}_{m'+n'}(\mu)-M^{(n)}_{m'}(\mu))-n'\beta_n(\nu,\mu)\right|\stackrel{n',m'}{<<}(n')^{-\theta},$$

which implies

$$\sup_{n\geq 1}\sup_{n',m'\geq 1}(n')^{-(1-\theta)}\sup_{1\leq\nu}E[(M^{(n)}_{m'+n'}(\nu)-M^{(n)}_{m'}(\nu))^2]<\infty, \tag{2.42}$$

Proof: A direct calculation leads to

$$\begin{aligned}C_{Mn}(\nu,\mu)&=E[(M^{(n)}_{m'+n'}(\nu)-M^{(n)}_{m'}(\nu))(M^{(n)}_{m'+n'}(\mu)-M^{(n)}_{m'}(\mu))\,|\,\mathcal{F}_{m'}]\\&=C^{(1)}_{Mn}(\nu,\mu)+C^{(2)}_{Mn}(\nu,\mu)+C^{(3)}_{Mn}(\nu,\mu),\end{aligned}$$

where

$$\begin{aligned}C^{(1)}_{Mn}(\nu,\mu)&=\sum_{k=m'+1}^{m'+n'}\widetilde{f}^{(n)}_{0,0}(\nu)\widetilde{f}^{(n)}_{0,0}(\mu)(\gamma_k-\sigma_k^4),\\C^{(2)}_{Mn}(\nu,\mu)&=\sum_{k=m'+1}^{m'+n'}\sum_{l=0}^{k-m'+1}\widetilde{f}^{(n)}_{l,0}(\nu)\widetilde{f}^{(n)}_{l,0}(\mu)\sigma_k^2\sigma_{k-l}^2,\\C^{(3)}_{Mn}(\nu,\mu)&=\sum_{k=m'+1}^{m'+n'}\sum_{l,l'=k-m'}^{\infty}\widetilde{f}^{(n)}_{l,0}(\nu)\widetilde{f}^{(n)}_{l',0}(\mu)\sigma_k^2\epsilon_{k-l}\epsilon_{k-l'},\end{aligned}$$

for $m',n'\geq 0$. First notice that

$$\sup_{n\geq 1}\sup_{1\leq\nu,\mu}|C^{(1)}_{Mn}(\nu,\mu)+C^{(2)}_{Mn}(\nu,\mu)-n'\beta_n(\nu,\mu)|\stackrel{n',m'}{\ll}(n')^{1-\theta},$$

see (A.10). Therefore, (2.39) follows if we show that

$$\sup_{n\geq 1}E\left[\sup_{1\leq\nu,\mu}|C^{(3)}_{Mn}(\nu,\mu)|\right]\stackrel{n',m'}{\ll}(n')^{1-\theta}.$$

Recall that $c_2=\sup_{k\geq 1}E(\epsilon_k^2)<\infty$ and assume w.l.o.g. $c_2=1$ in what follows. The Cauchy–Schwarz inequality yields

$$\begin{aligned}&|(n')^{-1}C^{(3)}_{Mn}(\nu,\mu)|\\&\quad\leq\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\left|\sum_{l=k-m'}^{\infty}\widetilde{f}^{(n)}_{l,0}(\nu)\epsilon_{k-l}\right|\left|\sum_{l'=k-m'}^{\infty}\widetilde{f}^{(n)}_{l',0}(\mu)\sigma_k^2\epsilon_{k-l'}\right|\\&\quad\leq\sqrt{\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\left(\sum_{l=k-m'}^{\infty}|\widetilde{f}^{(n)}_{l,0}(\nu)|\epsilon_{k-l}\right)^2}\sqrt{\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\left(\sum_{l=k-m'}^{\infty}|\widetilde{f}^{(n)}_{l,0}(\mu)|\epsilon_{k-l'}\right)^2}.\end{aligned}$$

Using $\|\cdot\|_{\ell_1} \le \|\cdot\|_{\ell_4}$ and Jensen's inequality, we obtain

$$\sqrt{\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\left(\sum_{l=k-m'}^{\infty}|\widetilde{f}_{l,0}^{(n)}(\nu)|\epsilon_{k-l}\right)^2} \le \sqrt{\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\left(\sum_{l=k-m'}^{\infty}[\widetilde{f}_{l,0}^{(n)}(\nu)]^4\epsilon_{k-l}^4\right)^{1/2}}$$
$$\le \sqrt{\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\left(\sum_{l=k-m'}^{\infty}\sup_{1\le\nu}[\widetilde{f}_{l,0}^{(n)}(\nu)]^4\epsilon_{k-l}^4\right)^{1/2}}$$
$$\le \left(\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\sum_{l=k-m'}^{\infty}\sup_{1\le\nu}[\widetilde{f}_{l,0}^{(n)}(\nu)]^4\epsilon_{k-l}^4\right)^{1/4},$$

where the upper bound does not depend on ν. Hence,

$$E\sup_{1\le\nu,\mu}|(n')^{-1}C_{Mn}^{(3)}(\nu,\mu)| \le E\sqrt{\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\sum_{l=k-m'}^{\infty}\sup_{1\le\nu}[\widetilde{f}_{l,0}^{(n)}(\nu)]^4\epsilon_{k-l}^4}$$
$$\le \sqrt{\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\sum_{l=k-m'}^{\infty}\sup_{1\le\nu}[\widetilde{f}_{l,0}^{(n)}(\nu)]^4\gamma_{k-l}}$$

Using $|\widetilde{f}_{l,0}^{(n)}(\nu)| = O(\|\mathbf{v}_n^{(\nu)}\|_{\ell_1}\|\mathbf{v}_n^{(\mu)}\|_{\ell_1} l^{-3/4-\theta/2})$, uniformly in n, $\sup_{k\ge1}\gamma_k < \infty$ and the elementary fact that $\sum_{k=m'}^{m'+n'}\sum_{l=k-m'+1} l^{-3-2\theta} = O((n')^{1-2\theta})$, (2.39) follows. The above arguments also imply that

$$\sup_{n\ge1} E(n')^{-1+\theta}|C_{Mn}(\nu,\mu)| \overset{n',m'}{\ll} 1,$$

which in turn implies (2.40). To verify (2.41) one first conditions on $\mathcal{F}_{m'}$ and then argues similarly in order to estimate $EC_{Mn}^{(3)}(\nu,\mu)$. Observe that with $c_{24} = \max\{\sup_{k\ge1}(\sigma_k^2)^3, \sup_{k\ge1}\gamma_k^4\sup_{k\ge1}\sigma_k^2\}$

$$E(n')^{-1}C_{Mn}^{(3)}(\nu,\mu) \le c_{24}\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\sum_{l=0}^{\infty}\left(\sup_{1\le\nu}|f_{l,0}^{(n)}(\nu)|\right)^2$$
$$\le c_{24}\sqrt{\frac{1}{n'}\sum_{k=m'+1}^{m'+n'}\sum_{l=0}^{\infty}\left(\sup_{1\le\nu}|f_{l,0}^{(n)}(\nu)|\right)^4}$$
$$\overset{n',m'}{\ll} (n')^{-\theta},$$

using $\|\cdot\|_{\ell_1} \le \|\cdot\|_{\ell_4}$ and Jensen's inequality, which verifies (2.41) and in turn (2.42). *Q.E.D.*

Introduce for $m', n' \ge 1$ and each coordinate $1 \le \nu \le L_n$ the partial sums $D_{n',m'}^{(n)}(\mathbf{v}_n^{(\nu)}, \mathbf{w}_n^{(\nu)}) = \sum_{i=m'+1}^{m'+n'}\xi_i^{(n)}(\mathbf{v}_n^{(\nu)}, \mathbf{w}_n^{(\nu)})$ and denote the appropriately scaled versions by

$$T_{n',m'}^{(n)}(\nu) = \frac{D_{n',m'}^{(n)}(\mathbf{v}_n^{(\nu)}, \mathbf{w}_n^{(\nu)})}{\sqrt{L_n n'}}, \qquad m', n' \ge 1, \tag{2.43}$$

for $1 \le \nu \le L_n$. The corresponding martingale approximations are given by

$$M_{n',m'}^{(n)}(\nu) = \frac{M_{n'+m'}^{(n)}(\mathbf{v}_n^{(\nu)}, \mathbf{w}_n^{(\nu)}) - M_{m'}^{(n)}(\mathbf{v}_n^{(\nu)}, \mathbf{w}_n^{(\nu)})}{\sqrt{L_n n'}}, \qquad m', n' \ge 1. \tag{2.44}$$

We need to study the approximation error,

$$R_{n',m'}^{(n)}(\nu) = T_{n',m'}^{(n)}(\nu) - M_{n',m'}^{(n)}(\nu), \qquad n', m' \ge 1.$$

The next result improves upon [Lemma 2][11] by showing that, firstly, the error is of order $(n')^{-\theta}$ in terms of the L_1–norm when conditioning on the past, and, secondly, that the result is uniform over ℓ_1–bounded weighting vectors.

Lemma 2.3 *We have*

$$\left\| \sup_{1 \le \nu} E\left((R_{n',m'}^{(n)}(\nu))^2 \,|\, \mathcal{F}_{m'} \right) \right\|_1 \overset{n',m'}{\ll} L_n^{-1}(n')^{-\theta}.$$

Proof: Consider, as in [11], the decomposition

$$R_{n',m'}^{(n)}(\nu) = Q_{n',m'}^{(n)}(\nu) + P_{n',m'}^{(n)}(\nu) + O_{n',m'}^{(n)}(\nu),$$

where

$$Q_{n',m'}^{(n)}(\nu) = \frac{1}{\sqrt{L_n n'}} \sum_{i=0}^{n'-1} \sum_{l=0}^{n'-i-1} \widetilde{f}_{l,i+1}^{(n)} (\sigma_{m'-n'-i}^2 \mathbf{1}_{\{l=0\}} - \epsilon_{m'+n'-i}\epsilon_{m'+n'-l}), \tag{2.45}$$

$$P_{n',m'}^{(n)}(\nu) = \frac{1}{\sqrt{L_n n'}} \sum_{i=0}^{\infty} \sum_{l=0}^{\infty} (\widetilde{f}_{l,i+1}^{(n)}(\nu) - \widetilde{f}_{l,i+n'+1}^{(n)}(\nu))(\epsilon_{m'-i}\epsilon_{m'-i-l} - \sigma_{m'-l}^2 \mathbf{1}_{\{l=0\}}), \tag{2.46}$$

$$O_{n',m'}^{(n)}(\nu) = -\frac{1}{\sqrt{L_n n'}} \sum_{i=0}^{n'-1} \sum_{k=n'}^{\infty} \widetilde{f}_{k-i,i+1}^{(n)}(\nu) \epsilon_{m'+n'-k}\epsilon_{m'+n'-i}. \tag{2.47}$$

$P_{n',m'}^{(n)}(\nu)$ is the projection of $R_{n',m'}^{(n)}(\nu)$ onto the subspace spanned by $\{\epsilon_r\epsilon_s - \sigma_r^2 \mathbf{1}_{\{r=s\}} : -\infty < r, s \le m'\}$ and therefore $\mathcal{F}_{m'}$–measurable. Hence with $e_{m',i,l} = \epsilon_{m'-i}\epsilon_{m'-i-l} - \sigma_{m'-l}^2 \mathbf{1}_{\{l=0\}}$

$$\begin{aligned}
L_n n' E\left((P_{n',m'}^{(n)}(\nu))^2 \,|\, \mathcal{F}_{m'} \right) &= \left(\sum_{i=0}^{\infty} \sum_{l=0}^{\infty} (\widetilde{f}_{l,i+1}^{(n)}(\nu) - \widetilde{f}_{l,i+n'+1}^{(n)}(\nu)) e_{m',i,l} \right)^2 \\
&\le \left(\sum_{i=1}^{\infty} \sum_{l=0}^{\infty} |\widetilde{f}_{l,i+1}^{(n)}(\nu) - \widetilde{f}_{l,i+n'+1}^{(n)}(\nu)||e_{m',i,l}| \right)^2 \\
&\le \sum_{i=1}^{\infty} \sum_{l=0}^{\infty} (\widetilde{f}_{l,i+1}^{(n)}(\nu) - \widetilde{f}_{l,i+n'+1}^{(n)}(\nu))^2 e_{m',i,l}^2 \\
&\le \sup_{k \ge 1} \gamma_k^2 \sup_{1 \le \nu} \sum_{i=0}^{\infty} \sum_{l=0}^{\infty} (\widetilde{f}_{l,i+1}^{(n)}(\nu) - \widetilde{f}_{l,i+n'+1}^{(n)}(\nu))^2,
\end{aligned}$$

a.s., such that due to (A.4)

$$E \sup_{1 \le \nu} E\left((P_{n',m'}^{(n)}(\nu))^2 \,|\, \mathcal{F}_{m'} \right) \overset{m',n'}{<<} L_n^{-1}(n')^{-\theta}, a.s.$$

$Q_{n',m'}^{(n)}(\nu)$ is the projection of $R_{n',m'}^{(n)}(\nu)$ onto the subspace spanned by $\{\epsilon_r\epsilon_s - \sigma_r^2 \mathbf{1}_{\{r=s\}} : m' < r, s \le m' + n'\}$ and thus independent from $\mathcal{F}_{m'}$, such that $E\left((Q_{n',m'}^{(n)}(\nu))^2 \,|\, \mathcal{F}_{m'} \right) = E(Q_{n',m'}^{(n)}(\nu))^2 \overset{m',n'}{<<} \sup_{k\ge1} \gamma_k (L_n n')^{-1} \sum_{i=1}^{n'} \sum_{l=0}^{n'-i} (\widetilde{f}_{l,i}^{(n)}(\nu))^2 \overset{m',n'}{<<} (n')^{-\theta}$, uniformly in $1 \le \nu$. Last, by Fatou

$$\begin{aligned}
&(L_n n')E\left((O_{n',m'}^{(n)}(\nu))^2 \,|\, \mathcal{F}_{m'} \right) \\
&\quad \le \lim_{N\to\infty} \sum_{i,i'=0}^{n'-1} \sum_{k,k'=n'}^{N} \widetilde{f}_{k-i,i+1}^{(n)}(\nu) \widetilde{f}_{k-i',i'+1}^{(n)}(\nu) \epsilon_{m'+n'-k}\epsilon_{m'+n'-k'} E(\epsilon_{m'+n'-i}\epsilon_{m'+n'-i'}) \\
&\quad \le \sup_{k\ge1} \sigma_k^2 \lim_{N\to\infty} \sum_{i=0}^{n'-1} \sum_{k,k'=n'}^{N} |\widetilde{f}_{k-i,i+1}^{(n)}(\nu) \widetilde{f}_{k'-i,i+1}^{(n)}(\nu) \epsilon_{m'+n'-k}\epsilon_{m'+n'-k'}| \\
&\quad \le \sup_{k\ge1} \sigma_k^2 \lim_{N\to\infty} \sum_{i=0}^{n'-1} \sqrt{ \sum_{k,k'=n'}^{N} [\widetilde{f}_{k-i,i+1}^{(n)}(\nu)]^2 [\widetilde{f}_{k'-i,i+1}^{(n)}(\nu)]^2 \epsilon_{m'+n'-k}^2 \epsilon_{m'+n'-k'}^2 } \\
&\quad \le \sup_{k\ge1} \sigma_k^2 \lim_{N\to\infty} \sum_{i=0}^{n'-1} \sum_{k=n'}^{N} [\widetilde{f}_{k-i,i+1}^{(n)}(\nu)]^2 \epsilon_{m'+n'-k}^2
\end{aligned}$$

a.s., where we estimated the ℓ_1–norm by the ℓ_2–norm. Hence,

$$\begin{aligned}
E \sup_{1\le\nu} E\left((O_{n',m'}^{(n)}(\nu))^2 \,|\, \mathcal{F}_{m'} \right) &\le (L_n n')^{-1} \sup_{k\ge1} (\sigma_k^2)^2 \sum_{i=0}^{n'-1} \sum_{l=1}^{\infty} \sup_{1\le\nu} [\widetilde{f}_{l,i}^{(n)}(\nu)]^2 \\
&\overset{m',n'}{\ll} L_n^{-1}(n')^{-\theta},
\end{aligned}$$

by virtue of (A.6). This completes the proof. *Q.E.D.*

Proof of Theorem 2.2: For a sequence of conditional covariance operators $C_n(\bullet|\mathcal{A}) = E[(\bullet, X_n)X_n \,|\, \mathcal{A}]$ with $X_n = (X_{nj})_j$, $E(X_n) = 0$, $n \ge 1$, say, we have convergence in the operator semi-norm, defined as $\|T\| = \sup_{f:\|f\|=1} |(f, Tf)|$ for an operator T acting on ℓ_2, to some unconditional covariance operator $C(\bullet) = E[(\bullet, Z)Z]$, $Z = (Z_j)_j$, $E(Z) = 0$, in expectation, if

$$E \sup_{f\in\ell_2:\|f\|=1} |(f, C_n(f|\mathcal{A}) - C(f))| = E \sup_{f\in\ell_2:\|f\|=1} \left| \sum_{j,k=1}^{\infty} f_j f_k [E(X_{nj}X_{nk}|\mathcal{A}) - E(Z_jZ_k)] \right|$$

converges to 0, as $n \to \infty$. Define the ℓ_2–valued random elements

$$M_{n',m'}^{(n)} = \left(M_{n',m'}^{(n)}(\nu) \right)_{\nu=1}^{\infty}, \qquad T_{n',m'}^{(n)} = \left(T_{n',m'}^{(n)}(\nu) \right)_{\nu=1}^{\infty},$$

where $T_{n',m'}^{(n)}(\nu) = 0$ and $M_{n',m'}^{(n)}(\nu) = T_{n',m'}^{(n)}(\nu)$, for $\nu > L_n$. Recall that

$$\mathcal{C}_{n',m'}^{(n)}(f|\mathcal{F}_m) = E[(f, T_{n',m'}^{(n)})T_{n',m'}^{(n)}|\mathcal{F}_{m'}], \qquad f \in \ell_2,$$

and let

$$\overline{\mathcal{C}}_{n',m'}^{(n)}(f|\mathcal{F}_m) = E[(f, M_{n',m'}^{(n)})M_{n',m'}^{(n)}|\mathcal{F}_{m'}], \qquad f \in \ell_2,$$

be the conditional covariance operator associated to the martingale approximations. Obviously,

$$\sup_{n \geq 1} E\|\mathcal{C}_{n',m'}^{(n)}(\bullet|\mathcal{F}_{m'}) - \mathcal{C}^{(n)}(\bullet)\| \leq \Xi_{n',m'} + \Psi_{n',m'}$$

where

$$\begin{aligned}
\Xi_{n',m'} &= \sup_{n \geq 1} E\|\mathcal{C}_{n',m'}^{(n)}(\bullet|\mathcal{F}_{m'}) - \overline{\mathcal{C}}_{n',m'}^{(n)}(\bullet|\mathcal{F}_{m'})\|, \\
\Psi_{n',m'} &= \sup_{n \geq 1} E\|\overline{\mathcal{C}}_{n',m'}^{(n)}(\bullet|\mathcal{F}_{m'}) - \mathcal{C}^{(n)}(\bullet)\|.
\end{aligned}$$

We shall estimate both terms separately. To simplify notation, let

$$\begin{aligned}
C_{Tn}(\nu, \mu) &= E\left(T_{n',m'}^{(n)}(\nu)T_{n',m'}^{(n)}(\mu) \,|\, \mathcal{F}_{m'}\right), \\
C_{Mn}(\nu, \mu) &= E\left(M_{n',m'}^{(n)}(\nu)M_{n',m'}^{(n)}(\mu) \,|\, \mathcal{F}_{m'}\right),
\end{aligned}$$

for $1 \leq \nu, \mu \leq L_n$. To estimate $\Xi_{n',m'}$, we shall show that $|C_{Tn}(\nu, \mu) - C_{Mn}(\nu, \mu)$ is $O(L_n^{-1}(n')^{-\theta/2})$, uniformly in n, ν, μ. By an application of the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&E \sup_{1 \leq \nu, \mu} |C_{Tn}(\nu, \mu) - C_{Mn}(\nu, \mu)| \\
&= E \sup_{1 \leq \nu, \mu} \left| E\left(T_{n',m'}^{(n)}(\nu)T_{n',m'}^{(n)}(\mu) \,|\, \mathcal{F}_{m'}\right) - E\left(M_{n',m'}^{(n)}(\nu)M_{n',m'}^{(n)}(\mu) \,|\, \mathcal{F}_{m'}\right) \right| \\
&\leq E \sup_{1 \leq \nu, \mu} E\left(|T_{n',m'}^{(n)}(\nu) - M_{n',m'}^{(n)}(\nu)||T_{n',m'}^{(n)}(\mu)| \,\middle|\, \mathcal{F}_{m'} \right) \\
&\quad + E \sup_{1 \leq \nu, \mu} E\left(|M_{n',m'}^{(n)}(\nu)||T_{n',m'}^{(n)}(\mu) - M_{n',m'}^{(n)}(\mu)| \,\middle|\, \mathcal{F}_{m'} \right),
\end{aligned}$$

where

$$\begin{aligned}
&E \sup_{1 \leq \nu, \mu} E\left(|T_{n',m'}^{(n)}(\nu) - M_{n',m'}^{(n)}(\nu)||T_{n',m'}^{(n)}(\mu)| \,|\, \mathcal{F}_{m'} \right) \\
&\quad \leq E \sup_{1 \leq \nu} \sqrt{E\left((T_{n',m'}^{(n)}(\nu) - M_{n',m'}^{(n)}(\nu))^2 \,|\, \mathcal{F}_{m'} \right)} \sqrt{\sup_{1 \leq \nu} E|T_{n',m'}^{(n)}(\nu)|^2} \\
&\quad \overset{n',m'}{\ll} L_n^{-1}(n')^{-\theta/2},
\end{aligned}$$

since $T_{n',m'}^{(n)}(\mu)$ is independent from $\mathcal{F}_{m'}$ and the decomposition $T_{n',m'}^{(n)}(\mu) = M_{n',m'}^{(n)}(\mu) + (T_{n',m'}^{(n)}(\mu) - M_{n',m'}^{(n)}(\mu))$ leads to $\sup_{n \geq 1} \sup_{1 \leq \nu} E|T_{n',m'}^{(n)}(\nu)|^2 < \infty$, by virtue of (2.42) and

Lemma 2.3. Further,

$$
\begin{aligned}
&E\left(|M^{(n)}_{n',m'}(\nu)||T^{(n)}_{n',m'}(\mu) - M^{(n)}_{n',m'}(\mu)| \,\big|\, \mathcal{F}_{m'}\right)\\
&\quad \le \sqrt{E\left((T^{(n)}_{n',m'}(\mu) - M^{(n)}_{n',m'}(\mu))^2 \,|\, \mathcal{F}_{m'}\right)}\sqrt{E\left(|M^{(n)}_{n',m'}(\nu)|^2 \,|\, \mathcal{F}_{m'}\right)}\\
&\quad \overset{n',m'}{\ll} L_n^{-1}(n')^{-\theta/2},
\end{aligned}
$$

by (2.40) and Lemma 2.3. uniformly in $\nu, \mu = 1, \dots, L_n$. Consequently,

$$
\sup_{n\ge1} L_n E \sup_{1\le \nu,\mu} |C_{Tn}(\nu,\mu) - C_{Mn}(\nu,\mu)| \overset{n',m'}{\ll} (n')^{-\theta/2}
$$

Hence, using the inequality $\sum_{\nu=1}^{L_n} |f_\nu| \le L_n^{1/2}\left(\sum_{\nu=1}^{L_n} f_\nu^2\right)^{1/2}$, we obtain

$$
\begin{aligned}
\Xi_{n',m'} &\le \sup_{n\ge1} E\|\mathcal{C}^{(n)}_{n',m'}(\bullet|\mathcal{F}_{m'}) - \overline{\mathcal{C}}^{(n)}_{n',m'}(\bullet|\mathcal{F}_{m'})\|\\
&= \sup_{n\ge1} E \sup_{f:\|f\|_{\ell_2}=1} \left|\sum_{\nu=1}^{L_n}\sum_{\mu=1}^{L_n} f_\nu f_\mu (C^T_{\nu\mu} - C^M_{\nu\mu})\right|\\
&\le \sup_{n\ge1} E \sup_{1\le\nu,\mu} |C^T_{\nu\mu} - C^M_{\nu\mu}| \sup_{f:\|f\|_{\ell_2}=1} \sum_{\nu=1}^{L_n}\sum_{\mu=1}^{L_n} |f_\nu f_\mu|\\
&\overset{n',m'}{\ll} L_n^{-1}(n')^{-\theta/2}\left(\sum_{\nu=1}^{L_n}|f_\nu|\right)^2 \overset{n',m'}{\ll} (n')^{-\theta/2},
\end{aligned}
$$

By Lemma 2.2 and the scaling of the martingale approximations, $M^{(n)}_{n',m'}(\nu)$, $1 \le \nu \le L_n$, by the factor $(L_n n')^{-1/2}$,

$$
\sup_{n\ge1} L_n \max_{1\le\nu,\mu\le L_n} \|E[M^{(n)}_{n',m'}(\nu)M^{(n)}_{n',m'}(\mu)|\mathcal{F}_{m'}] - L_n^{-1}\beta_n^2(\nu,\mu)\|_{L_1} \overset{n',m'}{\ll} (n')^{-\theta}.
$$

Therefore

$$
\begin{aligned}
\Psi_{n',m'} &= \sup_{n\ge1} E\|\overline{\mathcal{C}}^{(n)}_{n',m'}(\bullet|\mathcal{F}_m) - \mathcal{C}^{(n)}(\bullet)\|\\
&\le \sup_{n\ge1} E \sup_{\|f\|_{\ell_2}\le1} \left|\sum_{\nu=1}^{L_n}\sum_{\mu=1}^{L_n} f_\nu f_\mu (C^M_{\nu\mu} - L_n^{-1}\beta_n^2(\nu,\mu))\right|\\
&\le \sup_{n\ge1} E \sup_{\|f\|_{\ell_2}=1} \sup_{1\le\nu,\mu} |C^M_{\nu\mu} - L_n^{-1}\beta_n^2(\nu,\mu)| \left(\sum_{\nu=1}^{L_n}|f_\nu|\right)^2\\
&\overset{n',m'}{\ll} \sup_{n\ge1} L_n^{-1}(n')^{-\theta} \sup_{\|f\|_{\ell_2}=1} L_n \sum_{\nu=1}^{L_n} f_\nu^2 \overset{n',m'}{\ll} (n')^{-\theta}.
\end{aligned}
$$

Q.E.D.

PROOF OF THEOREM 2.3: By virtue of Lemma 2.1, Equation (2.20), we have the representations

$$D_{nt} = \sum_{k \le t} \left(\xi_k^{(n)}(j) \right)_{j=1}^{L_n}, \quad \mathcal{D}_n(t) = \frac{1}{\sqrt{nL_n}} \sum_{k \le \lfloor nt \rfloor} \left(\xi_k^{(n)}(j) \right)_{j=1}^{L_n},$$

and therefore we check conditions (I) – (III) of [15] discussed above for $\zeta_k = L_n^{-1/2} \xi_k^{(n)}$, cf. (2.37). The summands can be seen as attaining values in the Euclidean space $\mathbb{R}^{L_n}$ of finite (but increasing in n) dimension L_n or as random elements taking values in the infinite dimensional Hilbert space ℓ_2.

To show (I) observe that by the C_r–inequality, for each $1 \le j \le L_n$,

$$\begin{aligned} E|\xi_k^{(n)}(j)|^{2+\delta} &\le E(|Y_{nk}(\mathbf{v}_n^{(j)})Y_{nk}(\mathbf{w}_n^{(j)})| + E|Y_{nk}(\mathbf{v}_n^{(j)})Y_{nk}(\mathbf{w}_n^{(j)})|)^{2+\delta} \\ &\le 2^{3+\delta} E|Y_{nk}(\mathbf{v}_n^{(j)})Y_{nk}(\mathbf{w}_n^{(j)})|^{2+\delta} \\ &\le 2^{3+\delta} \sqrt{E|Y_{nk}(\mathbf{v}_n^{(j)})|^{4+2\delta}} \sqrt{E|Y_{nk}(\mathbf{w}_n^{(j)})|^{4+2\delta}}. \end{aligned}$$

Repeating the arguments of [11, p. 343], we obtain for $\delta' \in (0, 2)$ with $\chi = \delta'/2$

$$\begin{aligned} E|Y_{nk}(\mathbf{v}_n^{(j)})|^{4+\delta'} \le& \sup_{k' \ge 0} E|\epsilon_{k'}|^{4+\delta'} \sum_{l=0}^{\infty} |c_{n\ell}^{(v_j)}|^{2(2+\chi)} \\ &+ \sup_{k' \ge 0} E(\epsilon_{k'}^2) \left\{ \sup_{k' \ge 1} E(\epsilon_{k'}^2) \right\}^{1+\chi} \sum_{\ell=0}^{\infty} |c_{n\ell}^{(v_j)}|^2 \left\{ \sum_{\ell'=0}^{\infty} |c_{n\ell}^{(v_j)}|^2 \right\}^{1+\chi}, \end{aligned}$$

uniformly in $k \ge 1$. But $\sup_{1 \le j} |c_{n\ell}^{(v_j)}| \le \|\mathbf{v}_n^{(j)}\|_{\ell_1} \sup_{1 \le \nu} |c_{n\ell}^{(\nu)}|$ and $\sup_{1 \le \nu} |c_{n\ell}^{(\nu)}| \le (l \vee 1)^{-3/4-\theta/2}$, due to Assumption (A), imply $\sup_{1 \le j} \sum_{\ell=0}^{\infty} |c_{n\ell}^{(v_j)}|^2 < \infty$ and, in turn, $\sum_{l=0}^{\infty} |c_{n\ell}^{(v_j)}|^{2(2+\chi)} < \infty$. Noting that the above bounds hold uniformly in k and n, we obtain

$$\sup_{n \ge 1} \sup_{k \ge 1} \max_{1 \le j \le L_n} E|\xi_k^{(n)}(j)|^{2+\delta} < \infty.$$

By virtue of Jensen's inequality, we may now conclude that

$$E\|L_n^{-1/2} \xi_k^{(n)}\|_{\ell_2}^{2+\delta} = E \left(\frac{1}{L_n} \sum_{j=1}^{L_n} [\xi_k^{(n)}(j)]^2 \right)^{1+\delta/2} \le L_n^{-1} \sum_{j=1}^{L_n} E|\xi_k^{(n)}(j)|^{2+\delta} < \infty,$$

which establishes (I).

Introduce the partial sums

$$S_{n',m'}^{(n)} = \frac{1}{\sqrt{n' L_n}} \sum_{k=m'+1}^{m'+n'} \xi_k^{(n)}, \qquad n', m' \ge 1. \tag{2.48}$$

Condition (II) can be shown as follows. Denote the coordinates of $S_{n',m'}^{(n)}$ by $S_{n',m'}^{(n)}(j)$ and notice that they are given by $S_{n',m'}^{(n)}(j) = \sum_{k=m'+1}^{m'+n'} L_n^{-1/2} \xi_k^{(n)}(j)$. Denote the corresponding martingale approximations by $M_{n',m'}^{(n)}$ and $M_{n',m'}^{(n)}(j)$, respectively, and let $R_{n',m'}^{(n)} = S_{n',m'}^{(n)} -$

$M^{(n)}_{n',m'}$ be the remainder with coordinates $R^{(n)}_{n',m'}(j)$, $1 \le j \le L_n$, cf. the preparations above. Clearly, the martingale property implies $E(S^{(n)}_{n',m'}(j)|\mathcal{F}_{m'}) = E(R^{(n)}_{n',m'}(j)|\mathcal{F}_{m'})$, $1 \le j \le L_n$. Lemma 2.3 asserts that

$$\sup_{n\ge1} L_n E\left[\sup_{1\le j} E\left((R^{(n)}_{n',m'}(j))^2 \,|\, \mathcal{F}_{m'}\right)\right] \stackrel{n',m'}{\ll} (n')^{-\theta},$$

such that two applications of Jensen's inequality lead to

$$\begin{aligned}
\sup_{n\ge1} E\left|E(S^{(n)}_{n',m'} \,|\, \mathcal{F}_{m'})\right|_{\ell_2} &\le \sup_{n\ge1}\sqrt{\sum_{j=1}^{L_n} E\left[E(R^{(n)}_{n',m'}(j) \,|\, \mathcal{F}_{m'})\right]^2} \\
&\le \sup_{n\ge1}\sqrt{\sum_{j=1}^{L_n} E\left[E\left((R^{(n)}_{n',m'}(j))^2 \,|\, \mathcal{F}_{m'}\right)\right]} \\
&\le \sup_{n\ge1}\sqrt{L_n E\left[\sup_{1\le j} E\left((R^{(n)}_{n',m'}(j))^2 \,|\, \mathcal{F}_{m'}\right)\right]} \\
&\stackrel{n',m'}{\ll} (n')^{-\theta/2},
\end{aligned}$$

which shows (II). Condition (III) follows from Theorem 2.2.

Consequently, we may conclude that we may redefine all processes on a rich enough probability space where a Brownian motion $B_n(t) = (B_n(t)_j)_j$ with covariance operator $\mathcal{C}^{(n)}$, i.e. with covariances $E(B_n(t)_j B_n(t)_k) = L_n^{-1}\beta_n^2(j,k)$, exists, such that for constants $\lambda_n > 0$ and $C_n < \infty$

$$\|L_n^{-1/2} D_{n,t}(\mathbf{v}_n, \mathbf{w}_n) - B_n(t)\|_{\mathbb{R}^{L_n}} \le C_n t^{1/2-\lambda_n}, \qquad \text{for all } t \ge 0,$$

a.s.. Therefore

$$\sup_{t\in[0,1]} \|\mathcal{D}_n(t) - \overline{B}_n(\lfloor nt\rfloor/n)\|_{\mathbb{R}^{L_n}} \le C_n n^{-\lambda_n},$$

a.s., for the $[0,1]$–version $\overline{B}_n$ of B_n, which implies assertions (i) and (ii).

To show (iii) recall that the vector 1–norm of $\mathbb{R}^{L_n}$ can be bounded by $L_n^{-1/2}\| \bullet \|_{\mathbb{R}^{L_n}_n}$, such that

$$\sum_{j=1}^{L_n} |\mathcal{D}_{nj}(t) - \overline{B}_n\left(\lfloor nt\rfloor/n\right)_j| \le L_n^{1/2}\|\mathcal{D}_n(t) - \overline{B}_n(\lfloor nt\rfloor/n)\|_{\ell_2} = o(L_n^{1/2}),$$

as $n \to \infty$, a.s. Further, using $|x_j| \le \sqrt{\sum_j x_j^2}$, we have

$$\sup_{t\in[0,1]} |\mathcal{D}_n(t; \mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}) - \overline{B}_n(\lfloor nt\rfloor/n)_j| \le \sup_{t\in[0,1]} \|\mathcal{D}_n(t) - \overline{B}_n(\lfloor nt\rfloor/n)\|_{\ell_2} = o(1),$$

as $n \to \infty$, a.s., for $j = 1, \ldots, L_n$.

It remains to prove (iv). We may argue as in [17] to obtain

$$\begin{aligned}\mathbf{v}_n^{(\rho)\prime}(\widehat{\boldsymbol{\Sigma}}_{nmk}-\boldsymbol{\Sigma}_{nmk})\mathbf{w}_m^{(\sigma)} &= \sum_{i\le k}[(\mathbf{v}_n^{(\rho)\prime}\mathbf{Y}_{ni})(\mathbf{w}_m^{(\sigma)\prime}\mathbf{Y}_{mi}) - E((\mathbf{v}_n^{(\rho)\prime}\mathbf{Y}_{ni})(\mathbf{w}_m^{(\sigma)\prime}\mathbf{Y}_{mi}))]\\ &= \sum_{i\le k}[Y_{ni}(\mathbf{v}_n^{(\rho)})'Y_{mi}(\mathbf{w}_m^{(\sigma)}) - E(Y_{ni}(\mathbf{v}_n^{(\rho)})'Y_{mi}(\mathbf{w}_m^{(\sigma)}))]\end{aligned}$$

where

$$Y_{ni}(\mathbf{v}_n^{(\rho)}) = \sum_{j=0}^{\infty}\sum_{\nu=1}^{d_n} v_{n\nu}^{(\rho)}c_{nj}^{(\nu)}\epsilon_{i-j} \quad \text{and} \quad Y_{mi}(\mathbf{w}_m^{(\sigma)}) = \sum_{j=0}^{\infty}\sum_{\mu=1}^{d_n} w_{m\mu}^{(\sigma)}c_{mj}^{(\mu)}\epsilon_{i-j},$$

for $\rho = 1,\dots,L_n$ and $\sigma = 1,\dots,L_m$, $n,m\ge 1$. Therefore, for $k\ge 1$, we obtain the representation

$$\begin{aligned}&D_k(\{(\mathbf{v}_n^{(\rho)},\mathbf{w}_n^{(\rho)})\})\\ &\quad= \sum_{i\le k}\sum_{n,m}\sum_{\rho,\sigma}\lambda_n\lambda_m\mu_\rho\mu_\sigma[\mathbf{Y}_{ni}(\mathbf{v}_n^{(\rho)})'\mathbf{Y}_{mi}(\mathbf{w}_m^{(\sigma)}) - E(\mathbf{Y}_{ni}(\mathbf{v}_n^{(\rho)})'\mathbf{Y}_{mi}(\mathbf{w}_m^{(\sigma)}))]\\ &\quad= \sum_{i\le k}[Y_i(\{c_j\})Y_i(\{d_j\}) - E(Y_i(\{c_j\})Y_i(\{d_j\}))]\end{aligned}$$

for the linear processes $Y_i(\{c_j\}) = \sum_{j=0}^{\infty}c_j\epsilon_{i-j}$ and $Y_i(\{d_j\}) = \sum_{j=0}^{\infty}d_j\epsilon_{i-j}$ with coefficients

$$\begin{aligned}c_j &= \sum_{n=1}^{\infty}\lambda_n\sum_{\rho=1}^{L_n}\mu_\rho\sum_{\nu=1}^{d_n}v_{n\nu}^{(\rho)}c_{nj}^{(\nu)}, \qquad j\ge 0,\\ d_j &= \sum_{n=1}^{\infty}\lambda_n\sum_{\rho=1}^{L_n}\mu_\rho\sum_{\nu=1}^{d_n}w_{n\nu}^{(\rho)}c_{nj}^{(\nu)}, \qquad j\ge 0.\end{aligned}$$

Hence the result follows from [11]. *Q.E.D.*

Proof of Remark 2.1: By Theorem 2.1, we may and will assume that, on the same probability space,

$$|\mathcal{D}_n(t;\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)}) - \alpha_n(\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)})W_n(\lfloor nt\rfloor/n;\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)})| = o(1),$$

as $n\to\infty$, for L_n standard Brownian motions $W_n(\lfloor nt\rfloor/n;\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)})$, by virtue of Theorem 2.1. But then, since $\mathcal{D}_{nj}(t) = L_n^{-1/2}\mathcal{D}_n(t;\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)})$,

$$\begin{aligned}&\sup_{t\in[0,1]}\left|\overline{B}_n(\lfloor nt\rfloor/n)_j - L_n^{-1/2}\alpha_n(\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)})W_n(\lfloor nt\rfloor/n;\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)})\right|\\ &\le \sup_{t\in[0,1]}|\overline{B}_n(\lfloor nt\rfloor/n)_j - \mathcal{D}_{nj}(t)|\\ &+ \sup_{t\in[0,1]}L_n^{-1/2}|\mathcal{D}_n(t;\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)}) - \alpha_n(\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)})W_n(\lfloor nt\rfloor/n;\mathbf{v}_n^{(j)},\mathbf{w}_n^{(j)})|\\ &= o(1),\end{aligned}$$

for each $j = 1,\dots,L_n$, which verifies the remark. *Q.E.D.*

Proof of Theorem 2.4: Observe that the conditions (I)-(III) of [15, Theorem 1] hold in the Hilbert space ℓ_2 as well, since for any $\mathbf{x} \in \mathbb{R}^{d_n}$ the Euclidean vector norm coincides with the ℓ_2–norm. Therefore, we obtain the a.s. strong approximation

$$\left\| \sum_{i=1}^k L_n^{-1/2} \xi_i^{(n)} - B_n(k) \right\|_{\ell_2} = \epsilon_{nk} \sqrt{k \log\log k},$$

as $k \to \infty$, for sequences $\varepsilon_{nk} = o(1)$, $k \to \infty$, a.s., $n \geq 1$. Put $N = \lceil n \log\log n \rceil$. Let $\eta_n \downarrow 0$ be given. Then for each $n \in \mathbb{N}$ we may find $\delta_n \in \mathbb{N}$ such that $P(\max_{\delta_n \leq k'} \epsilon_{nk'} > \varepsilon) \leq \eta_n$. Hence $\max_{\delta_n \leq k'} \epsilon_{nk'} = o_P(1)$, as $n \to \infty$. Now we may conclude that for $\delta_n \leq k \leq n$

$$\left| \sum_{i=1}^k L_n^{-1/2} \xi_i^{(n)} - B_n(k) \right| \leq \max_{\delta_n \leq k'} \varepsilon_{nk'} \sqrt{N},$$

such that

$$\max_{\delta_n \leq k \leq n} \frac{1}{\sqrt{N}} \left| \sum_{i=1}^k L_n^{-1/2} \xi_i^{(n)} - B_n(k) \right| = o_P(1),$$

as $n \to \infty$, which verifies (2.31)–(2.33). *Q.E.D.*

3. ASYMPTOTICS FOR THE TRACE NORM

The trace plays an important role in multivariate analysis and also arises when studying shrinkage estimation. Before providing the large sample approximation by a Brownian motion, we shall briefly review its relation to several matrix norms.

3.1. *The trace and related matrix norms*

There are various matrix norms that can be used to measure the size of (covariance) matrices. Here we shall use the trace norm defined as the ℓ_1-norm of the eigenvalues $\lambda_i(\mathbf{A})$ of a d_n-dimensional matrix $\mathbf{A}$,

$$\|\mathbf{A}\|_{tr} = \sum_i |\lambda_i(\mathbf{A})|.$$

Also notice that the trace norm is a linear mapping on the subspace of non-negative definite matrices and satisfies $\|\mathbf{A}\|_{tr} = \operatorname{tr}(\mathbf{A})$ for any covariance matrix $\mathbf{A}$. It induces the Frobenius norm via $\|\mathbf{A}\|_F^2 = \operatorname{tr}(\mathbf{A}\mathbf{A}')$. Further, it is worth mentioning that the trace norm is also related to the Frobenius norm via the fact

$$\|\mathbf{A}^{1/2}\|_F^2 = \sum_i \lambda_i(\mathbf{A}) = \|\mathbf{A}\|_{tr}.$$

In this way, our results formulated in terms of (scaled) trace norms can be interpreted in terms of (scaled) squared Frobenius norms of square roots, too.

There is a third interesting direct link to another family of norms, namely the Schatten-p norms $\|\mathbf{A}\|_{S,p}$, $p \geq 0$, of a $n \times m$ matrix $\mathbf{A}$ of rank r, which is defined as the ℓ_p-norm of its (non-negative) singular values $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_r(\mathbf{A}) > \sigma_{r+1}(\mathbf{A}) = \cdots = \sigma_n(\mathbf{A}) = 0$, i.e. of the eigenvalues of $|\mathbf{A}| = (\mathbf{A}\mathbf{A}')^{1/2}$, i.e.

$$\|\mathbf{A}\|_{S,p}^p = \sum_i \sigma_i(\mathbf{A})^p.$$

The Schatten-1 norm $\|\mathbf{A}\|_{S,1}$ is also called nuclear norm. Since $\widehat{\mathbf{\Sigma}}_n = \mathbf{A}\mathbf{A}'$, if $\mathbf{A} = \mathcal{Y}_n/\sqrt{n}$, such that $\lambda_i(\widehat{\mathbf{\Sigma}}_n) = \sigma_i(\mathcal{Y}_n/\sqrt{n})^2$, we have the identity

$$\|\widehat{\mathbf{\Sigma}}_n\|_{tr} = \sum_i \lambda_i(\widehat{\mathbf{\Sigma}}_n) = \|\mathcal{Y}_n/\sqrt{n}\|_{S,2}^2.$$

between the trace norm of the sample covariance matrix and the Schatten-2 norm of the scaled data matrix.

For a sequence $\{\mathbf{A}_n\}$ of matrices of growing dimension $d_n \times d_n$ it makes sense to attach a scalar weight depending on the dimension to a given norm, such that *simple* matrices such as the identity matrix receive bounded norms. Having in mind that the squared Frobenius norm of $\mathbf{A}_n$ is the trace of $\mathbf{A}_n\mathbf{A}_n'$, it is natural to attach a scalar weight $f(d_n)$ to the trace operator leading to the scalar weight $f(d_n)^{1/2}$ for the Frobenius norm. As proposed by [13], one may select $f(d_n)$ such that $\operatorname{tr}(\mathbf{A}^*)f(d_n) = 1$ for some *simple* benchmark matrix $\mathbf{A}^*$ such as the d_n–dimensional identity matrix I_n. Since $\operatorname{tr}(I_{d_n}) = d_n$, we choose $f(d_n) = d_n^{-1}$ and therefore define the *scaled trace operator* by

$$\operatorname{tr}^*(\mathbf{A}) = d_n^{-1} \sum_{i=1}^{d_n} \lambda_i(\mathbf{A})$$

for a square matrix $\mathbf{A} = (a_{ij})_{i,j}$ of dimension $d_n \times d_n$. The scaled trace operator induces the *scaled trace norm*

$$\|\mathbf{A}\|_{tr}^* = d_n^{-1}\|\mathbf{A}\|_{tr}$$

for a square matrix $\mathbf{A}$, which is given by $\|\mathbf{A}\|_{tr}^* = d_n^{-1}\operatorname{tr}(\mathbf{A})$ for a covariance matrix and averages the (modulus) of the eigenvalues, and the *scaled Frobenius matrix norm* given by

$$\|\mathbf{A}\|_F^{*2} = \operatorname{tr}^*(\mathbf{A}\mathbf{A}') = d_n^{-1}\|\mathbf{A}\|_F^2. \tag{3.1}$$

3.2. *Trace asymptotics*

Let us now turn to the trace asymptotics. If the dimension is fixed, it is well known that the eigenvalues of a sample covariance matrix, and thus their sum as well, have convergence rate $O_P(n^{-1/2})$ and are asymptotically normal, see [9] and [10]. For the high-dimensional case, the situation is more involved. The sample covariance matrix is not consistent w.r.t. to the Frobenius norm, even in the presence of a dimension reducing factor model, see Remark 3.1.

The following result provides a large sample normal approximation for the scaled trace norm of $\widehat{\mathbf{\Sigma}}_n$ for arbitrarily growing dimension d_n when properly normalized. The result also shows that the trace norm has convergence rate

$$\big|\|\widehat{\mathbf{\Sigma}}_n\|_{tr} - \|\mathbf{\Sigma}_n\|_{tr}\big| = O_P(n^{-1/2}d_n),$$

as $n \to \infty$.

Introduce for $t \in [0,1]$

$$\widehat{\boldsymbol{\Sigma}}_n(t) = \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} \mathbf{Y}_{ni} \mathbf{Y}_{ni}', \qquad \boldsymbol{\Sigma}_n(t) = E(\widehat{\boldsymbol{\Sigma}}_n(t)), \tag{3.2}$$

and notice that

$$\widehat{\boldsymbol{\Sigma}}_n = \widehat{\boldsymbol{\Sigma}}_n(1) \qquad \text{and} \qquad \boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_n(1).$$

We are interested in studying the scaled trace norm process

$$\mathcal{T}_n(t) = \sqrt{n} \left(\| \widehat{\boldsymbol{\Sigma}}_n(t) \|_{tr}^* - \| \boldsymbol{\Sigma}_n(t) \|_{tr}^* \right), \qquad t \in [0,1]. \tag{3.3}$$

THEOREM 3.1 *Let $\mathbf{Y}_{ni}$, $i = 1, \dots, n$, be a vector time series following model (1.1) and satisfying Assumption (A). If (2.26) holds, then under the construction of Theorem 2.3,*

$$\left| \mathcal{T}_n(t) - d_n^{-1/2} \sum_{j=1}^{d_n} \overline{B}_n(\lfloor nt \rfloor / n)_j \right| = o(1),$$

as $n \to \infty$, a.s. Here $\overline{B}_n$ denotes the $[0,1]$–version of the Brownian motion B_n arising in Theorem 2.3, when choosing the d_n pairs $(\mathbf{v}_n^{(j)}, \mathbf{w}_n^{(j)}) = (\mathbf{e}_j, \mathbf{e}_j)$, $j = 1, \dots, d_n$, where $\mathbf{e}_j$ denotes the jth unit vector, and satisfies properties (ii) and (iii) of Theorem 2.3.

Suppose that, in addition to the assumptions of Theorem 3.1, $\mathbf{Y}_{n1}, \dots, \mathbf{Y}_{nn}$ is strictly stationary. Since the weighting vectors used in Theorem 3.1 are the first d_n unit vectors, the covariance of $\overline{B}_n(1)_i$ and $\overline{B}_n(1)_j$, which is associated to the asymptotic covariance of $\mathcal{D}_{ni}(1)_i$ and $\mathcal{D}_{nj}(1)_j$, is given by $d_n^{-1} \beta_n^2(i,j)$ where $\beta_n^2(i,j) = \beta_n^2(\mathbf{e}_i, \mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_j)$, $i, j = 1, \dots, d_n$, cf. (A.9). We have the asymptotic representations

$$\begin{aligned}
\beta_n^2(i,j) &= \operatorname{Cov}\left(\sqrt{n} \mathbf{e}_i' (\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n) \mathbf{e}_i, \sqrt{n} \mathbf{e}_j' (\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n) \mathbf{e}_j \right) + o(1) \\
&= \operatorname{Cov}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n [(Y_{nk}^{(i)})^2 - \sigma^2], \frac{1}{\sqrt{n}} \sum_{k=1}^n [(Y_{nk}^{(j)})^2 - \sigma^2] \right) + o(1) \\
&= \frac{1}{n} \sum_{k,k'=1}^n E\left((Y_{nk}^{(i)})^2 - \sigma^2 \right) \left((Y_{nk'}^{(j)})^2 - \sigma^2 \right) + o(1).
\end{aligned}$$

Therefore, we may express $\beta_n^2(i,j)$ as a long–run variance parameter,

$$\beta_n^2(i,j) = \gamma_n^{(i,j)}(0) + 2 \sum_{\tau=1}^{n-1} \frac{n-\tau}{n} \gamma_n^{(i,j)}(\tau), \tag{3.4}$$

where

$$\gamma_n^{(i,j)}(\tau) = \operatorname{Cov}\left((Y_{n0}^{(i)})^2, (Y_{n\tau}^{(j)})^2 \right), \qquad \tau = 0, \dots, n-1, \tag{3.5}$$

is the lag τ cross–covariance of the series $\{(Y_{nk}^{(i)})^2 : k \geq 0\}$ and $\{(Y_{nk}^{(j)})^2 : k \geq 0\}$, $i, j = 1, \dots, d_n$. Those cross-covariances can be estimated by

$$\widehat{\gamma}_n^{(i,j)}(\tau) = \frac{1}{n} \sum_{k=1}^{n-\tau} [Y_k^2(\mathbf{e}_i) - \widehat{\mu}_n(i)][Y_{k+\tau}^2(\mathbf{e}_j) - \widehat{\mu}_n(j)]$$

with $\widehat{\mu}_n(i) = n^{-1} \sum_{k=1}^n Y_k^2(\mathbf{e}_i)$, where $Y_k(\mathbf{e}_i) = \mathbf{e}_i' \mathbf{Y}_{nk}$, for $k = 1, \dots, n$, $i = 1, \dots, d_n$, $n \geq 1$. The associated estimator for $\beta_n^2(i, j)$ is then given by

$$\widehat{\beta}_n^2(i, j) = \widehat{\boldsymbol{\gamma}}_n^{(i,j)}(0) + 2 \sum_{\tau=1}^{m} w_{m\tau} \widehat{\boldsymbol{\gamma}}_n^{(i,j)}(\tau) \ ,$$

where $m = m_n$ is a sequence of lag truncation constants and $\{w_{mh}\}$ a sequence of window weights typically defined by a kernel function w (e.g. a Bartlett kernel) via $w_{m\tau} = w(\tau/b_m)$, for some bandwidth parameter $b = b_m$.

By Theorem 3.1,

$$\operatorname{Var}(\mathcal{T}_n(1)) = \sigma_{tr,n}^2 + o(1), \qquad \sigma_{tr,n}^2 = \frac{1}{d_n} \sum_{j,k=1}^{d_n} \operatorname{Cov}(\overline{B}_n(1)_j, \overline{B}_n(1)_k),$$

a.s., and, using the canonical estimator

$$\widehat{\sigma}_{tr,n}^2 = \frac{1}{d_n} \sum_{j,k=1}^{d_n} \widehat{\beta}_n^2(i, j),$$

an asymptotic confidence interval with nominal coverage probability $1 - \alpha$, $\alpha \in (0, 1)$, for $\|\boldsymbol{\Sigma}_n\|_{tr}^*$ is given by

$$\|\widehat{\boldsymbol{\Sigma}}_n\|_{tr}^* \pm \Phi^{-1}(1 - \alpha/2) \widehat{\sigma}_{tr,n}.$$

Lemma 3.1 *Assume $w_{m\tau} \to 1$ as $m \to \infty$, for all $\tau \in \mathbb{Z}$, and $0 \leq w_{m\tau} \leq W < \infty$, for some constant W, for all $m \geq 1$, $\tau \in \mathbb{Z}$. Further, suppose that $c_{nj}^{(\nu)} = c_j^{(\nu)}$, $\nu \in \mathbb{N}$, satisfy the decay condition*

$$\sup_{1 \leq \nu} |c_j^{(\nu)}| \ll (j \vee 1)^{-(1+\delta)}$$

for some $\delta > 0$, and ϵ_k are i.i.d. with $E(\epsilon_1^8) < \infty$. If $m = m_n \to \infty$ with $m^2/n = o(1)$, as $n \to \infty$, then

$$\lim_{n \to \infty} E|\widehat{\eta}_n^2 - \eta_n^2| = 0 \quad \text{and} \quad \lim_{n \to \infty} E|\widehat{\sigma}_{n,tr}^2 - \sigma_{n,tr}^2| = 0.$$

Remark 3.1 *It is worth comparing our result with the following result obtained by [6] for a dimension-reducing factor model: Suppose that the generic random vector $\mathbf{Y}_n = (Y_1, \dots, Y_{d_n})'$ satisfies a factor model*

$$\mathbf{Y}_n = \mathbf{B}_n \mathbf{f} + \boldsymbol{\epsilon}$$

with $K = K(d_n) \leq d_n$ observable factors $\mathbf{f} = (f_1, \dots, f_K)'$, errors $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_{d_n})'$ and a $d_n \times K$ factor loading matrix $\mathbf{B}_n$. Then the sample covariance matrix of an i.i.d. sample $(\mathbf{Y}_1, \mathbf{f}_1), \dots, (\mathbf{Y}_n, \mathbf{f}_n)$ has the convergence rate $O_P(n^{-1/2} d_n K)$,

$$\|\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n\|_F = O_P(n^{-1/2} d_n K),$$

if $E\|\mathbf{Y}_n\|_F^2$, $\max_i E(f_i^4)$ and $\max_i E(\epsilon_i^4)$ are bounded, see [6, Theorem 1]. This means, compared to the rate for fixed dimension, the Frobenius norm is inflated by the factor $d_n K$.

3.3. *Proofs*

PROOF OF THEOREM 3.1: Clearly, $\widehat{\boldsymbol{\Sigma}}_k(1)$ is p.s.d. for all $k \in \mathbb{N}$ and thus $\widehat{\boldsymbol{\Sigma}}_n(t) = \frac{\lfloor nt \rfloor}{n} \widehat{\boldsymbol{\Sigma}}_{\lfloor nt \rfloor}$ as well. The fact that $\operatorname{tr}(\mathbf{A}) = \sum_i \mathbf{e}_i' \mathbf{A} \mathbf{e}_i$ leads to

$$\|\widehat{\boldsymbol{\Sigma}}_n(t)\|_{tr} - \|\boldsymbol{\Sigma}_n(t)\|_{tr} = \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_n(t)) - \operatorname{tr}(\boldsymbol{\Sigma}_n(t)) = \sum_{j=1}^{d_n} \mathbf{e}_j' (\widehat{\boldsymbol{\Sigma}}_n(t) - \boldsymbol{\Sigma}_n(t)) \mathbf{e}_j.$$

Let $\mathcal{D}_n = (\mathcal{D}_{nj})_{j=1}^{d_n}$ with $\mathcal{D}_{nj}(t) = d_n^{-1/2} \mathcal{D}_n(t; \mathbf{e}_j, \mathbf{e}_j)$ for $j = 1, \dots, d_n$. We shall apply Theorem 2.3 with $L_n = d_n$. Since $\| \bullet \|_{tr}^* = d_n^{-1} \operatorname{tr}(\bullet)$, we have

$$\mathcal{T}_n(t) = \frac{1}{\sqrt{d_n}} \sum_{j=1}^{d_n} \mathcal{D}_{nj}(t).$$

Now we can conclude that the process

$$\mathcal{E}_n(t) = \sqrt{n}(\|\widehat{\boldsymbol{\Sigma}}_n(t)\|_{tr}^* - \|\boldsymbol{\Sigma}_n(t)\|_{tr}^*) - \sum_{j=1}^{d_n} d_n^{-1/2} \overline{B}_n(\lfloor nt \rfloor / n)_j$$

satisfies, when redefined on a new probability space,

$$\begin{aligned}
|\mathcal{E}_n(t)| &= \frac{1}{\sqrt{d_n}} \left| \sum_{j=1}^{d_n} \left(\mathcal{D}_{nj}(t) - \overline{B}_n(\lfloor nt \rfloor / n)_j \right) \right| \\
&\leq \frac{1}{\sqrt{d_n}} \sum_{i=1}^{d_n} \left| \mathcal{D}_{nj}(t) - \overline{B}_n(\lfloor nt \rfloor / n)_j \right| \\
&= o(1),
\end{aligned}$$

as $n \to \infty$, a.s., by Theorem 2.3 (iv). *Q.E.D.*

PROOF OF LEMMA 3.1: The proof follows easily from [17, Theorem 4.4]. *Q.E.D.*

4. SHRINKAGE ESTIMATION

Shrinkage is a well-established approach to regularize the sample variance–covariance matrix and we shall review in Section 4.1 the results obtained for high–dimensional settings. When shrinking towards the identity matrix in terms of a convex combination with the sample variance–covariance matrix, the optimal weight depends on the trace of the true variance–covariance matrix, which can be estimated canonically by the trace of the sample variance–covariance matrix. As a consequence, we can apply the results obtained in the previous section to obtain large sample approximations for shrinkage variance–covariance matrix estimators. Recall that the approximations deal with the norm of the difference between partial sums and a Brownian motion, both attaining values in a vector space. In order to compare covariance matrices, we shall work with the following pseudometric: Define

$$\Delta_n(\mathbf{A}_n, \mathbf{B}_n; \mathbf{v}_n, \mathbf{w}_n) = |\mathbf{v}_n'(\mathbf{A}_n - \mathbf{B}_n)\mathbf{w}_n|, \tag{4.1}$$

for (sequences of) matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ of dimension $d_n \times d_n$. Indeed, for fixed $\mathbf{v}_n, \mathbf{w}_n$ the mapping $(\mathbf{A}_n, \mathbf{B}_n) \mapsto \Delta_n(\mathbf{A}_n, \mathbf{B}_n) = \Delta_n(\mathbf{A}_n, \mathbf{B}_n; \mathbf{v}_n, \mathbf{w}_n)$ is symmetric, non-negative, semidefinite (i.e. $\mathbf{A}_n = \mathbf{B}_n$ implies $\Delta_n(\mathbf{A}_n, \mathbf{B}_n) = 0$) and satisfies the triangle inequality. Hence, (4.1) defines a pseudometric on the space of $(d_n \times d_n)$–dimensional matrices, for each n.

We establish three main results: For *regular* weighting vectors $\mathbf{v}_n, \mathbf{w}_n$ that are bounded away from orthogonality we establish a large sample approximation, which holds uniformly in the shrinkage weight and therefore also when using the common estimator for the optimal weight. Further, we compare the shrinkage estimator using the estimated optimal weight with an oracle estimator using the unknown optimal weight. In both cases, it turns out that the convergence rate of the estimated optimal shrinkage weight carries over to the shrinkage covariance estimator.

Lastly, we study the case of orthogonal and *nearly orthogonal* vectors. The latter case is of particular interest, since then one may place more unit vectors on the unit sphere corresponding to overcomplete bases as studied in areas such as dictionary learning.

4.1. *Shrinkage of covariance matrix estimators*

The results of the previous chapters show that, under general conditions, inference relying on ℓ_1-bounded inner products of high-dimensional series can be based on the sample covariance matrix, even if $d_n > n$ such that $\widehat{\mathbf{\Sigma}}_n$ is singular. However, from a statistical point of view, the use of this classical estimator is not recommended in such situations of high dimensionality: important criteria such as its mean-squared error or its condition number (defined to be the ratio of the largest to the smallest eigenvalue) deteriorate, and it is advisable to regularise $\widehat{\mathbf{\Sigma}}_n$ in order to improve its performance, both asymptotically and for finite sample sizes, with respect to these criteria. Obviously a particular interest lies in maximum-likelihood based approaches using an invertible estimator of $\mathbf{\Sigma}_n$ (as, e.g. in the semi-parametric approach of [7] on shrinkage estimation in multivariate hidden Markov models). One well-established possibility to regularise $\widehat{\mathbf{\Sigma}}_n$ without needing to impose any structural assumptions on $\mathbf{\Sigma}_n$, in particular avoiding sparsity, is the following approach of shrinkage: ([13], [16]) consider a shrinkage estimator defined by a linear (or convex) combination of $\widehat{\mathbf{\Sigma}}_n$ with a well-conditioned "target" matrix $\mathbf{T}_n$,

$$\mathbf{\Sigma}_n^s = \mathbf{\Sigma}_n^s(W_n) = (1 - W_n)\widehat{\mathbf{\Sigma}}_n \ + \ W_n \ \mathbf{T}_n \ , \tag{4.2}$$

where W_n are the "shrinkage weights" of this convex combination, to be chosen in an optimal way to minimise the mean-square error between $\mathbf{\Sigma}_n^s$ and $\mathbf{\Sigma}_n$ (see below). The role of the target $\mathbf{T}_n$ is, similar to ridge regression, to reduce a potentially large condition number of the high-dimensional variance-covariance matrix $\mathbf{\Sigma}_n$, by adding a highly regular ("well conditioned") matrix. A popular choice for the target is to take a multiple of the d_n−dimensional identity matrix I_n, i.e.

$$\mathbf{T}_n = \mu_n \; I_n \;, \tag{4.3}$$

with $\mu_n = \frac{1}{d_n} \operatorname{tr} \mathbf{\Sigma}_n$, in order to respect the scale of both matrices in the convex combination (4.2). This choice of the target reduces the dispersion of the eigenvalues of $\mathbf{\Sigma}_n$ around its "grand mean" $\mu_n = \frac{1}{d_n} \operatorname{tr} \mathbf{\Sigma}_n$, as large eigenvalues are pulled down towards μ_n and small eigenvalues are lifted up to μ_n (and in particular lifted up away from zero). Although a bias is introduced in estimating $\mathbf{\Sigma}_n$ by $\mathbf{\Sigma}_n^s$ compared to $\widehat{\mathbf{\Sigma}}_n$, the gain in variance reduction, in particular in high-dimensions, helps to considerably reduce the mean-square error in estimating $\mathbf{\Sigma}_n$.

In order to develop the correct asymptotic framework of the behaviour of large covariance matrices, the authors of [13] propose to use the *scaled Frobenius norm* given by (3.1) to measure the distance between two matrices of asymptotically growing dimension d_n, to be used also and in particular to define the mean-square error between $\mathbf{\Sigma}_n^s$ and $\mathbf{\Sigma}_n$ to become the expected normalised Frobenius loss $E[\|\mathbf{\Sigma}_n^s - \mathbf{\Sigma}_n\|_F^{*2}]$.

Furthermore, with this scaling,

$$\mu_n = \frac{1}{d_n} \operatorname{tr}(\mathbf{\Sigma}_n)$$

is the appropriate choice of the factor in front of the identity matrix I_n in the definition of the target $\mathbf{T}_n$ in equation (4.3).

In practice μ_n needs to be estimated from the trace of $\widehat{\mathbf{\Sigma}}_n$, i.e. by

$$\widehat{\mu}_n = \frac{1}{d_n} \operatorname{tr}(\widehat{\mathbf{\Sigma}}_n) \;.$$

Similarly, the theoretical shrinkage weight $0 \le W_n \le 1$ need to be replaced by its sample analog $\widehat{W}_n$. Thus, the fully data-driven expression for the shrinkage estimator of $\mathbf{\Sigma}$ writes as follows:

$$\mathbf{\Sigma}_n^s(\widehat{W}_n) = (1 - \widehat{W}_n)\widehat{\mathbf{\Sigma}}_n \; + \; \widehat{W}_n \widehat{\mu}_n I_n,$$

which shrinks the sample covariance matrix towards the (estimated) shrinkage target $\widehat{\mu}_n I_n$. It remains to optimally choose the shrinkage weights W_n (and its data-driven analogue $\widehat{W}_n$) with the purpose of balancing between a good fit and good regularisation. For this a prominent possibility is indeed to choose the shrinkage weights W_n such that the mean-squared error (MSE) between $\mathbf{\Sigma}_n^s$ and $\mathbf{\Sigma}_n$, is minimised:

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

which leads to the MSE-optimally shrunken matrix $\mathbf{\Sigma}_n^* = \mathbf{\Sigma}_n^s(W_n^*)$. A closed form solution ([13] or [16], Proposition 1) can be derived as

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

This choice leads to the interesting property that

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

showing the actual relative gain of the shrunken estimator compared to the classical unshrunken sample covariance, in terms of the mean-squared error. Moreover, it can be shown that this property continues to hold even if one replaces the in practice yet unknown optimal weights W_n^* by an estimator $\widehat{W}_n^*$ which is constructed by replacing the population quantities in numerator and denominator of W_n^* by sample analogs. Whereas the denominator can be essentially estimated by $\|\widehat{\mu}_n I_n - \widehat{\mathbf{\Sigma}}_n\|_F^{*2}$, it is slightly less straightforward to estimate the numerator $E[\|\widehat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n\|_F^2$: one possibility suggested by [16], and further developed by [17] for our set-up, is based on the estimation of the long-run variance α_n of $\widehat{\mathbf{\Sigma}}_n$.

Note that

$$E\|\widehat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n\|_F^{*2} = \frac{1}{nd_n}\sum_{i,j=1}^{d_n} \mathrm{Var}\,(\sqrt{n}\widehat{\mathbf{\Sigma}}_n(i,j)), \quad \widehat{\mathbf{\Sigma}}_n(i,j) = \frac{1}{n}\sum_{k=1}^{n}[Y_{nk}^{(i)}Y_{nk}^{(j)} - E(Y_{nk}^{(i)}Y_{nk}^{(j)})].$$

where, under stationarity, for $1 \le i, j \le d_n$,

$$\sigma_n^2(i,j) = \mathrm{Var}\,(\sqrt{n}\widehat{\mathbf{\Sigma}}_n(i,j)) = \Gamma_n^{(i,j)}(0) + 2\sum_{\tau=1}^{n-1}(n-\tau)\Gamma_n^{(i,j)}(\tau)$$

with $\Gamma_n^{(i,j)}(\tau) = \mathrm{Cov}\,(Y_{nk}^{(i)}Y_{nk}^{(j)}, Y_{n,k+\tau}^{(i)}Y_{n,k+\tau}^{(j)})$, $\tau = 0,\dots,n-1$, $n \ge 1$. A consistent estimator of the optimal weights W_n^* can now be obtained as follows. Let

$$\widehat{\Gamma}_n^{(i,j)}(\tau) = \frac{1}{n}\sum_{t=1}^{n-\tau}[Y_{nt}^{(i)}Y_{nt}^{(j)} - \widehat{\kappa}_n(i,j)]\;[Y_{n,t+\tau}^{(i)}Y_{n,t+\tau}^{(j)} - \widehat{\kappa}_n(i,j)]\;,$$

where $\widehat{\kappa}_n(i,j) = \frac{1}{n}\sum_{t=1}^{n} Y_{nt}^{(i)}Y_{nt}^{(j)}$, $1 \le i,j \le d_n$. Then, similar as in the previous section, the long-run variances $\sigma_n^2(i,j)$ can be estimated by

$$\widehat{\sigma}_n^2(i,j) = \widehat{\Gamma}_n^{(i,j)}(0) + 2\sum_{\tau=1}^{m} w_{m\tau}\widehat{\Gamma}_n^{(i,j)}(\tau)\;,$$

$1 \le i,j \le d_n$. Consistency of a more general version has been shown in [17, Equation (4.22)], under similar assumptions as stated in Lemma 3.1, for $d_n \to \infty$, as $n \to \infty$. We are led to the estimator

$$\widehat{W}_n^* = \frac{(nd_n)^{-1}\sum_{1\le i,j\le d_n}\widehat{\sigma}_n^2(i,j)}{\|\widehat{\mu}_n I_n - \widehat{\mathbf{\Sigma}}_n\|_F^{*2}}$$

for W_n^* also studied in depth in [16]. A rate of consistency in an asymptotic framework with growing dimensionality d_n can be achieved again following [16] for the specific shrinkage target $\mu_n I_n$, also considered in [7]. Let $0 < \gamma \le 2$ be such that $d_n^{2-\gamma}/n \to 0$ (i.e. the larger γ, the faster is d_n allowed to grow with n), and that, as $n \to \infty$,

$$d_n^{1-\gamma}\|\mu_n I_n - \mathbf{\Sigma}_n\|_F^{*2} \to c > 0\;.$$

Recalling that $\mu_n = d_n^{-1}\,\mathrm{tr}(\mathbf{\Sigma}_n)$, we observe that γ measures the closeness of the target to the true covariance matrix $\mathbf{\Sigma}_n$. Then [16] (Theorem 1) and ([7] (Theorem 1) show that

$$\frac{n}{d_n^{2-\gamma}}\;|\widehat{W}_n^* - W_n^*| = o_P(1)\;. \tag{4.4}$$

In order to apply the results of the previous sections onto the fully data-driven shrinkage estimator $\mathbf{\Sigma}_n^*(\widehat{W}_n^*)$, one needs to study the convergence of the estimated shrinkage weight normalised by $n^{1/2}$, as will become clear from the proof of Theorem 4.1 to be stated below.

We already observe here that (4.4) implies

$$n^{1/2}|\widehat{W}_n^* - W_n^*| = o_P(1), \tag{4.5}$$

if $d^{4-2\gamma}/n = O(1)$. Thus, for γ close to 2, the dimension d_n may even grow faster than n.

4.2. *Asymptotics for regular projections*

Our interest is now in deriving the asymptotics for bilinear forms based on the shrinkage estimator of the covariance matrix. We can and will assume that the uniformly ℓ_1–bounded weighting vectors $\mathbf{v}_n$ and $\mathbf{w}_n$ are ℓ_2–normed. It turns out that, due to the shrinkage target, the inner product $\mathbf{v}_n'\mathbf{w}_n$, i.e. the angle between the vectors $\mathbf{v}_n$ and $\mathbf{w}_n$, appears in the approximating Brownian functional. The inner product is bounded but may converge to 0, as n tends to ∞. The latter case requires special treatment and will be studied separately. We shall call a pair $(\mathbf{v}_n, \mathbf{w}_n)$ of projections *regular*, if it has uniformly bounded ℓ_1–norm and satisfies

$$\mathbf{v}_n'\mathbf{w}_n \geq \underline{c} > 0, \qquad \text{for all } n, \tag{4.6}$$

for some constant $\underline{c}$, i.e., if it is, in addition, bounded away from orthogonality.

Let $0 < W \leq 1$ be an arbitrary shrinkage weight and consider the associated shrinkage estimator

$$\widehat{\mathbf{\Sigma}}_n^s(W) = (1-W)\widehat{\mathbf{\Sigma}}_n + W\widehat{\mu}_n I_n.$$

Notice that $\widehat{\mathbf{\Sigma}}_n^s(W)$ estimates the unobservable shrunken variance matrix

$$\mathbf{\Sigma}_{n0}^s(W) = (1-W)\mathbf{\Sigma}_n + W\operatorname{tr}(\mathbf{\Sigma}_n)d_n^{-1}I_n.$$

Define for $0 < W \leq 1$

$$\mathcal{A}_n(W) = \sqrt{n}\mathbf{v}_n'(\widehat{\mathbf{\Sigma}}_n^s(W) - \mathbf{\Sigma}_{n0}^s(W))\mathbf{w}_n. \tag{4.7}$$

We shall apply the trace asymptotics obtained in Theorem 3.1, i.e.

$$\sqrt{n}(\|\widehat{\mathbf{\Sigma}}_n(t)\|_{tr}^* - \|\mathbf{\Sigma}_n\|_{tr}^*(t)) = d_n^{-1/2}\sum_{j=1}^{d_n}\overline{B}_n(\lfloor nt\rfloor/n)_j + o(1),$$

as $n \to \infty$, a.s. The variance of the approximating linear functional of the d_n–dimensional Brownian motion is given by

$$\sigma_{tr,n}^2 = \frac{1}{d_n}\sum_{i,j=1}^{d_n}\operatorname{Cov}(\overline{B}_n(1)_i, \overline{B}_n(1)_j) = \frac{1}{d_n^2}\sum_{i,j=1}^{d_n}\beta_n^2(i,j),$$

where $\beta_n^2(i,j)$ are long–run–variance parameters, see (3.4). Since, typically, long–run–variance parameters have positive limits, it is natural to assume that

$$\inf_{n\geq 1}\sigma_{tr,n}^2 \geq \underline{\sigma}_{tr}^2 > 0. \tag{4.8}$$

THEOREM 4.1 *Let $(\mathbf{v}_n, \mathbf{w}_n)$ be a regular pair of projections. Under the assumptions of Theorem 3.1 and condition (4.8) there exists on a new probability space, which carries an equivalent version of the vector time series, a Brownian motion $\overline{B}_n(t) = (\overline{B}_n(t)_j)_{j=0}^{d_n}$ on $[0,1]$, such that*

$$\sup_{0<W\leq 1} |\mathcal{A}_n(W) - \mathcal{B}_n(W)| = o(1), \tag{4.9}$$

as $n \to \infty$, a.s., where

$$\mathcal{B}_n(W) = (1-W)\overline{B}_n(1)_0 + W\mathbf{v}_n'\mathbf{w}_n d_n^{-1/2} \sum_{j=1}^{d_n} \overline{B}_n(1)_j, \tag{4.10}$$

The covariance structure of $\overline{B}_n(t)$ is given by

$$Var(\overline{B}_n(1)_0) = (d_n+1)^{-1}\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n), \quad Cov(\overline{B}_n(1)_0, \overline{B}_n(1)_i) = (d_n+1)^{-1}\beta_n^2(\mathbf{v}_n,\mathbf{w}_n,\mathbf{e}_i,\mathbf{e}_i),$$

for $i = 1, \ldots, d_n$, and

$$Cov(\overline{B}_n(1)_i, \overline{B}_n(1)_j) = (d_n+1)^{-1}\beta_n^2(i,j),$$

for $1 \leq i,j \leq d_n$. Especially, for any deterministic or random sequence of shrinkage weights W_n we have the large sample approximation for the corresponding shrinkage estimator

$$|\sqrt{n}\mathbf{v}_n'(\widehat{\mathbf{\Sigma}}_n^s(W_n) - \mathbf{\Sigma}_{n0}^s(W_n))\mathbf{w}_n - \mathcal{B}_n(W_n)| = o(1),$$

as $n \to \infty$, a.s..

Notice that

$$\begin{aligned}\mathrm{Var}\,(\mathcal{B}_n(W)) = {} & \frac{(1-W)^2}{d_n+1}\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n) + \frac{W^2(\mathbf{v}_n'\mathbf{w}_n)^2}{d_n(d_n+1)} \sum_{i,j=1}^{d_n} \beta_n^2(i,j) \\ & + 2\frac{(1-W)W\mathbf{v}_n'\mathbf{w}_n}{\sqrt{d_n(d_n+1)}} \sum_{j=1}^{d_n} \beta_n^2(\mathbf{v}_n,\mathbf{w}_n,\mathbf{e}_j,\mathbf{e}_j) + o(1),\end{aligned}$$

as $n \to \infty$. Hence, under assumption (4.6), the variance of the approximating Wiener process adressing the nonparametric part of the shrinkage estimator is of the order $O(d_n^{-1})$, whereas the variance of the term approximating the target is of the order $O(1)$. This is due to the fact that we need a (d_n+1)-dimensional Brownian motion (from which d_n coordinates are used to approximate the estimated target). This requires to scale *all* coordinates $(d_n+1)^{-1/2}$, cf. Theorem 2.3.

The following theorem resolves that issue by approximating the shrinkage estimator by two Brownian motions, one in dimension 1 for the nonparametric part and one in dimension d_n for the target. Those Brownian motions are constructed separately, such that, a priori, nothing can be said about their *exact* covariance structure. It turns out, however, that the covariances converge properly. We shall see that for this alternative construction the terms of the resulting variance-covariance decomposition are of the same order.

THEOREM 4.2 *Let $(\mathbf{v}_n, \mathbf{w}_n)$ be a regular pair of projections. Suppose that the underlying probability space $(\Omega, \mathcal{F}, P)$ is rich enough to carry, in addition to the vector time series $\{\mathbf{Y}_{ni} : 1 \le i \le n, n \ge 1\}$, a uniform random variable U_1. Then there exist, on $(\Omega, \mathcal{F}, P)$, a univariate Brownian motion $\{\overline{B}_n'(t)_0 : t \in [0,1]\}$ with mean zero and*

$$Cov(\overline{B}_n'(s)_0, \overline{B}_n'(t)_0) = \min(s,t)\alpha_n^2(\mathbf{v}_n; \mathbf{w}_n),$$

for $s, t \in [0,1]$, and a mean zero Brownian motion $\{(\overline{B}_n'(t)_j)_{j=1}^{d_n} : t \in [0,1]\}$ in dimension d_n with covariance function

$$Cov(\overline{B}_n'(s)_i, \overline{B}_n'(t)_j) = \min(s,t) d_n^{-1} \beta_n^2(i,j),$$

for $s, t \in [0,1]$ and $1 \le i, j \le d_n$, such that

$$\left| \sqrt{n}\mathbf{v}_n'(\widehat{\mathbf{\Sigma}}_n^s(W) - \mathbf{\Sigma}_{n_0}^s(W))\mathbf{w}_n - \mathcal{B}_n(W) \right| = o(1),$$

as $n \to \infty$, a.s., with $\mathcal{B}_n'(W) = (1-W)\overline{B}_n'(1)_0 + W\mathbf{v}_n'\mathbf{w}_n d_n^{-1/2} \sum_{j=1}^{d_n} \overline{B}_n'(1)_j$. Further,

$$\max_{1 \le j \le d_n} |Cov(\overline{B}_n'(1)_0, \overline{B}_n'(1)_j) - d_n^{-1/2}\beta_n^2(\mathbf{v}_n, \mathbf{w}_n, \mathbf{e}_j, \mathbf{e}_j) = o(1), \tag{4.11}$$

as $n \to \infty$.

Observe that

$$\begin{aligned}\mathrm{Var}\,(\mathcal{B}_n'(W)) &= (1-W)^2\alpha_n^2(\mathbf{v}_n, \mathbf{w}_n) + \frac{W^2(\mathbf{v}_n'\mathbf{w}_n)^2}{d_n^2} \sum_{i,j=1}^{d_n} \beta_n^2(i,j) \\ &\quad + 2\frac{(1-W)W\mathbf{v}_n'\mathbf{w}_n}{d_n} \sum_{j=1}^{d_n} \beta_n^2(\mathbf{v}_n, \mathbf{w}_n, \mathbf{e}_j, \mathbf{e}_j) + o(1),\end{aligned}$$

as $n \to \infty$, a.s., where all three terms are $O(1)$.

The above result shows that the nonparametric part, namely the sample covariance matrix $\widehat{\mathbf{\Sigma}}_n$, as well as the shrinkage target $\widehat{\mu}_n I_n$ contribute to the asymptotics. In this sense, shrinking with respect to the chosen scaled norms provides us with a large sample approximation that mimics the finite sample situation.

4.3. *Comparisons with oracle estimators*

Recall that an oracle estimator is an estimator that depends on quantities unknown to us such as the optimal shrinkage weight W_n^*. Of course, it is of interest to study the distance between the shrinkage estimator $\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*)$ with estimated optimal weight and the associated oracle using W_n^*. In particular, the question arises how the rate of convergence (4.4) affects the difference between the fully data adaptive estimator and an oracle.

The next theorem compares the shrinkage estimator $\widehat{\mathbf{\Sigma}}_n^s = \widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*)$ that uses the estimated optimal shrinkage weight $\widehat{W}_n^*$ and the *oracle estimator*

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

which shrinks the sample covariance matrix towards the target using the optimal shrinkage weight W_n^*, in terms of the pseudometric $\Delta_n(\bullet,\bullet;\mathbf{v}_n,\mathbf{w}_n)$ and thus considers the quantity

$$\Delta_n(\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*),\widehat{\mathbf{\Sigma}}_n^s(W_n^*);\mathbf{v}_n,\mathbf{w}_n) = \left|\mathbf{v}_n'(\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*)-\widehat{\mathbf{\Sigma}}_n^s(W_n^*))\mathbf{w}_n\right|.$$

The following result shows that even now the rate of convergence is equal to the rate of convergence of the estimator $\widehat{W}_n^*$.

Theorem 4.3 *Under the assumptions of Theorem 4.1 and the construction described there we have, on the new probability space,*

$$\Delta_n(\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*),\widehat{\mathbf{\Sigma}}_n^s(W_n^*);\mathbf{v}_n,\mathbf{w}_n)$$
$$= |\widehat{W}_n^*-W_n^*|\left(O(1)+O(\overline{B}_n(1)_0 n^{-1/2})+O\left(n^{-1/2}d_n^{-1/2}\sum_{i=1}^{d_n}\overline{B}_n(1)_i\right)+o(n^{-1/2}),\right)$$

as $n\to\infty$, a.s..

The next result investigates the difference between the shrinkage estimator and the oracle type estimator

$$\mathbf{\Sigma}_{n0}^s(W_n^*) = (1-W_n^*)\mathbf{\Sigma}_n + W_n^*\mathrm{tr}(\mathbf{\Sigma}_n)d_n^{-1}I_n$$

using the oracle shrinkage weight and assuming knowledge of $\mathbf{\Sigma}_n$, in terms of the pseudo-distance

$$\Delta_n(\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*),\mathbf{\Sigma}_{n0}^s(W_n^*);\mathbf{v}_n,\mathbf{w}_n) = \left|\mathbf{v}_n'[\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*)-\mathbf{\Sigma}_{n0}^s(W_n^*)]\mathbf{w}_n\right|.$$

Theorem 4.4 *Under the assumptions of Theorem 4.1 and the construction described there we have, on the new probability space,*

$$\Delta_n(\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*),\mathbf{\Sigma}_{n0}^s(W_n^*);\mathbf{v}_n,\mathbf{w}_n) = n^{-1/2}\mathcal{B}_n(\widehat{W}_n^*)+(W_n^*-\widehat{W}_n^*)O(1)+o(n^{-1/2}),$$

as $n\to\infty$, a.s..

The above result is remarkable in that it shows that it is optimal in the sense that $\Delta_n(\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*),\mathbf{\Sigma}_{n0}^s(W_n^*);\mathbf{v}_n,\mathbf{w}_n)$ inherits the rate of convergence from the estimator $\widehat{W}_n^*$ of the optimal shrinkage weight W_n^*, cf. (4.4).

4.4. *Nearly orthogonal projections*

Let $\mathbf{v}_n^{(i)}$, $i=1,\dots,L_n$, be unit vectors in $\mathbb{R}^{d_n}$ on which we may project $\mathbf{Y}_n$, e.g. in order to determine the best approximating direction. Recall that the true covariance between two projections $\mathbf{v}_n^{(i)\prime}\mathbf{Y}_n$ and $\mathbf{v}_n^{(j)\prime}\mathbf{Y}_n$ is

$$\mathrm{Cov}\,(\mathbf{v}_n^{(i)\prime}\mathbf{Y}_n,\mathbf{v}_n^{(j)\prime}\mathbf{Y}_n) = \mathbf{v}_n^{(i)\prime}\mathbf{\Sigma}_n\mathbf{v}_n^{(j)},$$

and the corresponding shrinkage estimator is

$$\widehat{\mathrm{Cov}}\,(\mathbf{v}_n^{(i)\prime}\mathbf{Y}_n,\mathbf{v}_n^{(j)\prime}\mathbf{Y}_n) = \mathbf{v}_n^{(i)\prime}\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*)\mathbf{v}_n^{(j)}.$$

Clearly, those covariances vanish for $i \neq j$, if the $\mathbf{v}_n^{(i)}$ are chosen as eigenvectors of $\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*)$, as in a classical principal component analysis (PCA) applied to the shrinkage covariance matrix estimator. But when analyzing high–dimensional data it is common to rely on procedures such as sparse PCA, see [19], which yield sparse principal components. Then analyzing the covariances of the projections $\mathbf{v}_n^{(i)\prime}\mathbf{Y}_n$ is of interest.

If $\mathcal{O}_{L_n} = \{\mathbf{v}_n^{(j)} : 1 \leq j \leq L_n\}$ is an orthogonal system and $L_n < d_n$, then $\mathcal{O}_{L_n}$ spans a L_n–dimensional subspace of $\mathbb{R}^{d_n}$. Of course, there are at most L_n orthogonal vectors, i.e. L_n cannot be larger than d_n. However, if one relaxes the orthogonality condition

$$\mathbf{v}_n^{(i)\prime}\mathbf{v}_n^{(j)} = 0, \qquad i \neq j,$$

then one can place much more unit vectors in the Euclidean space $\mathbb{R}^{d_n}$ in such a way that their pairwise angles are small. Indeed, [18] provides an elegant proof of the following Kabatjanskii-Levenstein bound.

THEOREM 4.5 (CHEAP VERSION OF THE KABATJANSKII-LEVENSTEIN BOUND, TAO 2013). *Let $\mathbf{x}_1, \dots, \mathbf{x}_m$ be unit vectors in $\mathbb{R}^{d_n}$ such that $|\mathbf{x}_i'\mathbf{x}_j| \leq A d_n^{-1/2}$ for some $1/2 \leq A \leq \frac{1}{2}\sqrt{d_n}$. Then we have $m \leq \left(\frac{Cd_n}{A^2}\right)^{CA^2}$ for some universal constant C.*

Theorem 4.5 motivates to study the case of *nearly orthogonal* weighting vectors, defined as a pair $(\mathbf{v}_n, \mathbf{w}_n) \in \mathcal{W} \times \mathcal{W}$ satisfying

$$\mathbf{v}_n'\mathbf{w}_n = o(1), \qquad n \to \infty, \tag{4.12}$$

Now the asymptotics of the shrinkage estimator is as follows:

THEOREM 4.6 *Let $\mathbf{v}_n$ and $\mathbf{w}_n$ be unit vectors satisfying the nearly orthogonal condition (4.12) and suppose that the conditions of Theorem 4.1 hold. Then*

$$|\sqrt{n}\mathbf{v}_n'(\widehat{\Sigma}_n^s(W) - \widehat{\mathbf{\Sigma}}_{n0}^s(W))\mathbf{w}_n - \mathcal{B}_n(W)| = o_P(1),$$

as $n \to \infty$, a.s., where

$$\mathcal{B}_n(W) = (1 - W)\overline{B}_n(1)_0 + W\mathbf{v}_n'\mathbf{w}_n d_n^{-1/2}\sum_{i=1}^{d_n}\overline{B}_n(1)_i.$$

Observe that for asymptotically orthogonal weighting vectors the term $W\mathbf{v}_n'\mathbf{w}_n d_n^{-1/2}\sum_{i=1}^{d_n}\overline{B}_n(1)_i$ corresponding to the (parametric) shrinkage target is $o_P(1)$ and thus vanishes asymptotically. In this situation, the nonparametric part dominates in large samples.

4.5. *Proofs*

PROOF OF THEOREM 4.1: First notice that (4.6) ensures that the second term in (4.10) does not converge to 0 in probability, since $(\mathbf{v}_n, \mathbf{w}_n)$ is a regular projection and condition (4.8) ensures that the Gaussian random variable $V_n = d_n^{-1/2}\sum_{j=1}^{d_n}\overline{B}_n(1)_j$ is not $o_P(1)$, since

$$\inf_{n\geq 1} P(|V_n| > \delta) \geq \inf_{n\geq 1} P(|V_n/\sigma_{tr,n}| > \delta/\underline{\sigma}) > 0$$

for any $\delta > 0$. Hence $W\mathbf{v}_n'\mathbf{w}_n V_n = o_P(1)$ iff. $\mathbf{v}_n'\mathbf{w}_n = o(1)$, which is excluded by (4.6).

We argue similarly as in the proof of Theorem 3.1: Put $\mathcal{D}_n = (\mathcal{D}_{nj})_{j=0}^{d_n}$ with $\mathcal{D}_{n0}(t) = L_n^{-1/2}\mathcal{D}_n(t;\mathbf{v}_n,\mathbf{w}_n)$ and $\mathcal{D}_{nj}(t) = L_n^{-1/2}\mathcal{D}_n(t,\mathbf{e}_j,\mathbf{e}_j)$, $j = 1,\dots,d_n$, where $L_n = d_n + 1$. Since the weighting vectors are uniformly ℓ_1-bounded, Theorem 2.3 yields, on a new probability space where a process equivalent to $\mathcal{D}_n$ can be defined and will be denoted again by $\mathcal{D}_n$, the existence of a Brownian motion $\{(\overline{B}_n(t))_{j=0}^{d_n} : t \geq 0\}$ as characterized in Theorem 4.1, such that

$$|\sqrt{n}(\mathbf{v}_n'\widehat{\boldsymbol{\Sigma}}_n\mathbf{w}_n - \mathbf{v}_n'\boldsymbol{\Sigma}_n\mathbf{w}_n) - \overline{B}_n(1)_0| = o(1),$$

and

$$\left|\sqrt{n}(\|\widehat{\boldsymbol{\Sigma}}_n\|_{tr}^* - \|\boldsymbol{\Sigma}_n\|_{tr}^*) - d_n^{-1/2}\sum_{i=1}^{d_n}\overline{B}_n(1)_i\right| = o(1),$$

as $n \to \infty$, a.s. Using these results, $\widehat{\mu}_n = \|\widehat{\boldsymbol{\Sigma}}_n\|_{tr} d_n^{-1}$ and the fact that $|\mathbf{v}_n'\mathbf{w}_n| \leq \|\mathbf{v}_n\|_{\ell_1}\|\mathbf{w}_n\|_{\ell_1} = O(1)$, we have for any $0 < W \leq 1$

$$\begin{aligned}
&|\mathcal{A}_n(W) - \mathcal{B}_n(W)| \\
&\quad= \Bigg|\sqrt{n}\mathbf{v}_n'(1-W)(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)\mathbf{w}_n - (1-W)\overline{B}_n(1)_0 \\
&\qquad + W\sqrt{n}\mathbf{v}_n'(\|\widehat{\boldsymbol{\Sigma}}_n\|_{tr}^* - \|\boldsymbol{\Sigma}_n\|_{tr}^*)\mathbf{w}_n - W\mathbf{v}_n'\mathbf{w}_n d_n^{-1/2}\sum_{i=1}^{d_n}\overline{B}_n(1)_i\Bigg| \\
&\quad\leq (1-W)|\sqrt{n}\mathbf{v}_n'(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)\mathbf{w}_n - \overline{B}_n(1)_0| \\
&\qquad + W\left|\sqrt{n}\mathbf{v}_n'\mathbf{w}_n(\|\widehat{\boldsymbol{\Sigma}}_n^*\|_{tr}^* - \|\boldsymbol{\Sigma}_n\|_{tr}^*) - \mathbf{v}_n'\mathbf{w}_n d_n^{-1/2}\sum_{i=1}^{d_n}\overline{B}_n(1)_i\right| \\
&\quad\leq (1-W)|\sqrt{n}\mathbf{v}_n'(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)\mathbf{w}_n - \overline{B}_n(1)_0| \\
&\qquad + W|\mathbf{v}_n'\mathbf{w}_n|\left|\sqrt{n}(\|\widehat{\boldsymbol{\Sigma}}_n^*\|_{tr}^* - \|\boldsymbol{\Sigma}_n\|_{tr}^*) - d_n^{-1/2}\sum_{i=1}^{d_n}\overline{B}_n(1)_i\right| \\
&\quad= o(1),
\end{aligned}$$

as $n \to \infty$, a.s., which shows (4.9). *Q.E.D.*

Proof of Theorem 4.2: By Theorem 2.1 there exist, on a new probability space $(\Omega', \mathcal{F}', P')$, an equivalent process $\{\widetilde{\mathbf{Y}}_{ni} : i \geq 1\} \stackrel{d}{=} \{\mathbf{Y}_{ni} : i \geq 1\}$ and a Brownian motion $\{\widetilde{B}_n(t)_0 : t \in [0,1]\}$ on $[0,1]$, such that

$$\sup_{t\in[0,1]}|\sqrt{n}\mathbf{v}_n'(\widetilde{\boldsymbol{\Sigma}}_n(t) - \boldsymbol{\Sigma}_n(t))\mathbf{w}_n - \widetilde{B}_n(t)_0| = o(1), \tag{4.13}$$

as $n \to \infty$, P'-a.s., where $\widetilde{\boldsymbol{\Sigma}}_n(t) = n^{-1}\sum_{i=1}^{\lfloor nt\rfloor}\widetilde{\mathbf{Y}}_{ni}\widetilde{\mathbf{Y}}_{ni}'$. By Billingsley's lemma, [2, Section 21, Lemma 2], there exists a Brownian motion $\{\overline{B}_n'(t)_0 : t \in [0,1]\} \stackrel{d}{=} \{\widetilde{B}_n(t)_0 : t \in [0,1]\}$ defined on the original probability space $(\Omega, \mathcal{F}, P)$, such that

$$|\sqrt{n}\mathbf{v}_n'(\widehat{\boldsymbol{\Sigma}}_n(t) - \boldsymbol{\Sigma}_n(t))\mathbf{w}_n - \overline{B}_n'(t)_0| = o(1), \tag{4.14}$$

as $n \to \infty$, a.s. (apply [2, Sec. 21, Lemma 2] with $\nu = \mathcal{L}(\sqrt{n}\mathbf{v}_n'(\widetilde{\mathbf{\Sigma}}_n(\bullet) - \mathbf{\Sigma}_n(\bullet))\mathbf{w}_n, \widetilde{B}_n(\bullet)_0)$ and $\sigma = \sqrt{n}\mathbf{v}_n'(\widetilde{\mathbf{\Sigma}}_n(\bullet) - \mathbf{\Sigma}_n(\bullet))\mathbf{w}_n$ to conclude the existence of $\tau =: \overline{B}_n'(\bullet)_0$, a function of σ and U_1, such that (4.14) holds).

Further, by Theorem 3.1, there exist, on a new probability space $(\Omega'', \mathcal{F}'', P'')$, an equivalent vector time series $\{\check{\mathbf{Y}}_{ni} : i \geq 1\} \stackrel{d}{=} \{\mathbf{Y}_{ni} : i \geq 1\}$, and a Brownian motion $\{(\widetilde{B}_n(t)_j)_{j=1}^{d_n} : t \in [0,1]\}$ on $[0,1]$ in dimension d_n characterized as in the theorem, such that

$$\left| \sqrt{n}(\|\check{\mathbf{\Sigma}}_n(t)\|_{tr}^* - \|\mathbf{\Sigma}_n(t)\|_{tr}^*) - d_n^{-1/2} \sum_{j=1}^{d_n} \widetilde{B}_n(1)_j \right| = o(1), \tag{4.15}$$

as $n \to \infty$, P''-a.s., where $\check{\mathbf{\Sigma}}_n(t) = n^{-1} \sum_{i=1}^{\lfloor nt \rfloor} \check{\mathbf{Y}}_{ni} \check{\mathbf{Y}}_{ni}'$. Again, an application of Billingsley's lemma shows the existence of a Brownian motion $\{(\overline{B}_n'(t)_j)_{j=1}^{d_n} : t \in [0,1]\} \stackrel{d}{=} \{(\widetilde{B}_n(t)_j)_{j=1}^{d_n} : t \in [0,1]\}$, such that

$$\left| \sqrt{n}(\|\widehat{\mathbf{\Sigma}}_n(t)\|_{tr}^* - \|\mathbf{\Sigma}_n(t)\|_{tr}^*) - d_n^{-1/2} \sum_{j=1}^{d_n} \overline{B}_n'(1)_j \right| = o(1), \tag{4.16}$$

as $n \to \infty$, a.s., i.e. on the original probability space.

A priori, we have no information on the exact second order structure of the two Brownian motions, but $\overline{B}_n'(1)_j$ is close to the associated process $\mathcal{D}_{nj}(t) = T_{n,0}^{(n)}(j)$, cf. (2.43), and to the corresponding martingale approximation $M_{\lfloor nt \rfloor,0}^{(n)}(j)$ defined in (2.44), which allows us to study the convergence of the covariances $\operatorname{Cov}(\overline{B}_n'(1)_0, \overline{B}_n'(1)_j) =< \overline{B}_n'(1)_0, \overline{B}_n'(1)_j >_{L_2}$, $j = 1, \dots, d_n$. First observe that

$$\max_{1 \leq j \leq d_n} \|\overline{B}_n'(1)_j - M_{n,0}^{(n)}(j)\|_{L_2} = o(1),$$

see [11, Lemma 2], [17], and Lemma 2.2, because of (4.13) and since $(\overline{B}_n'(t)_j)_{j=1}^{d_n}$ satisfies $\sup_{t \in [0,1]} \sum_{j=1}^{d_n} |\overline{B}_n'(t)_j - \mathcal{D}_{nj}(t)|^2 = o(1)$, as $n \to \infty$, a.s., see Theorem 2.3 (ii). Also notice that $\|\overline{B}_n'(1)_j\|_{L_2}$ and $\|M_{n,0}^{(n)}(j)\|_{L_2}$ are $O(1)$, uniformly in $j \geq 0$. Now use the decomposition

$$< X, Y >_{L_2} =< X', Y' >_{L_2} + < X', Y - Y' >_{L_2} + < X - X', Y >_{L_2}$$

for $X, Y, X', Y' \in L_2$ to conclude that

$$\begin{aligned} \operatorname{Cov}(\overline{B}_n'(1)_0, \overline{B}_n'(1)_j)) =&< M_{n,0}^{(n)}(0), M_{n,0}^{(n)}(j) >_{L_2} + < M_{n,0}^{(n)}(0), \overline{B}_n'(1)_j - M_{n,0}^{(n)}(j) >_{L_2} \\ &+ < \overline{B}_n'(1)_0 - M_{n,0}^{(n)}(0), M_{n,0}^{(n)}(j) >_{L_2}, \end{aligned}$$

where the last two terms are $o(1)$, uniformly in j. E.g.,

$$\max_{1 \leq j \leq d_n} | < \overline{B}_n'(1)_0 - M_{n,0}^{(n)}(0), M_{n,0}^{(n)}(j) >_{L_2} | \leq \sup_{1 \leq j} \|\overline{B}_n'(1)_0 - M_{n,0}^{(n)}(0)\|_{L_2} \|M_{n,0}^{(n)}(j)\|_{L_2} = o(1),$$

as $n \to \infty$. Combining these estimates with Lemma 2.2 yields

$$\max_{1 \leq j \leq d_n} |\operatorname{Cov}(\overline{B}_n'(1)_0, \overline{B}_n'(1)_j) - \operatorname{Cov}(M_{n,0}^{(n)}(0), M_{n,}^{(n)}(j))| \ll d_n^{-1} + o(1),$$

which establishes (4.11), since the covariances of the approximating martingales equal $d_n^{-1/2}\beta_n^2(\mathbf{v}_n, \mathbf{w}_n, \mathbf{e}_j, \mathbf{e}_j)+$ $o(1)$, as $n \to \infty$; the factor $d_n^{-1/2}$ is due to the additional scaling of $\mathcal{D}_{nj}(t)$ to approximate d_n bilinear forms by Theorem 2.3. *Q.E.D.*

PROOF OF THEOREM 4.3: Recall that, since $\mathbf{\Sigma}_n = \mathbf{C}_n \Lambda \mathbf{C}_n'$ and (1.2), the elements of $\mathbf{\Sigma}_n$ are uniformly bounded (in n), such that $|\mathbf{v}_n' \mathbf{\Sigma}_n \mathbf{w}_n| \le \|\mathbf{v}_n\|_{\ell_1} \|\mathbf{w}_n\|_{\ell_1} = O(1)$. This in turn implies that $\lambda_{\max}(\mathbf{\Sigma}_n) = O(1)$ and

$$\text{(4.17)} \qquad \operatorname{tr}(\Sigma_n) d_n^{-1} = O(1).$$

Put

$$\text{(4.18)} \qquad R_n(\widehat{W}_n^*, W_n^*) = \mathbf{v}_n' \mathbf{\Sigma}_{n0}^s(\widehat{W}_n^*)\mathbf{w}_n - \mathbf{v}_n' \mathbf{\Sigma}_{n0}^s(W_n^*)\mathbf{w}_n$$

and notice that

$$R_n(\widehat{W}_n^*, W_n^*) = (W_n^* - \widehat{W}_n^*)\mathbf{v}_n' \mathbf{\Sigma}_n \mathbf{w}_n + (\widehat{W}_n^* - W_n^*)\operatorname{tr}(\mathbf{\Sigma}_n) d_n^{-1} \mathbf{v}_n' \mathbf{w}_n.$$

Using (4.17) we obtain the bound

$$\text{(4.19)} \qquad R_n(\widehat{W}_n^*, W_n^*) = (\widehat{W}_n^* - W_n^*)O(1).$$

Observe that

$$\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*) - \widehat{\mathbf{\Sigma}}_n^s(W_n^*) = (W_n^* - \widehat{W}_n^*)\widehat{\mathbf{\Sigma}}_n + (\widehat{W}_n^* - W_n^*)\operatorname{tr}(\widehat{\mathbf{\Sigma}}_n) d_n^{-1} I_n$$

is equal to the difference

$$\mathbf{\Sigma}_{n0}^s(\widehat{W}_n^*) - \mathbf{\Sigma}_{n0}^s(W_n^*) = (W_n^* - \widehat{W}_n^*)\mathbf{\Sigma}_n + (\widehat{W}_n^* - W_n^*)\operatorname{tr}(\mathbf{\Sigma}_n) d_n^{-1} I_n$$

when replacing $\mathbf{\Sigma}_n$ by $\widehat{\mathbf{\Sigma}}_n$. We have

$$\begin{aligned}
&\widehat{\mathbf{\Sigma}}_n^s(\widehat{W}_n^*) - \widehat{\mathbf{\Sigma}}_n^s(W_n^*) \\
&\quad = (W_n^* - \widehat{W}_n^*)\mathbf{\Sigma}_n + (\widehat{W}_n^* - W_n^*)\operatorname{tr}(\mathbf{\Sigma}_n) d_n^{-1} I_n \\
&\quad + (W_n^* - \widehat{W}_n^*)(\widehat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n) + (\widehat{W}_n^* - W_n^*)(\operatorname{tr}(\widehat{\mathbf{\Sigma}}_n) - \operatorname{tr}(\mathbf{\Sigma}_n)) d_n^{-1} I_n
\end{aligned}$$

Using (4.19), we therefore obtain for the associated bilinear form

$$\begin{aligned}
&\mathbf{v}_n'(\widehat{\mathbf{\Sigma}}_n(\widehat{W}_n^*) - \widehat{\mathbf{\Sigma}}_n^s(W_n^*))\mathbf{w}_n \\
&\quad = R_n(\widehat{W}_n^*, W_n^*) \\
&\qquad + (W_n^* - \widehat{W}_n^*)\mathbf{v}_n'(\widehat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n)\mathbf{w}_n + (W_n^* - \widehat{W}_n^*)\mathbf{v}_n'\mathbf{w}_n(\|\widehat{\mathbf{\Sigma}}_n\|_{tr}^* - \|\mathbf{\Sigma}_n\|_{tr}^*) \\
&\quad = (\widehat{W}_n^* - W_n^*)O(1) \\
&\qquad + (W_n^* - \widehat{W}_n^*)\left(\overline{B}_n(1)_0 n^{-1/2} + o(n^{-1/2})\right) \\
&\qquad + (\widehat{W}_n^* - W_n^*)\left(n^{-1/2} d_n^{-1/2} \sum_{i=1}^{d_n} \overline{B}_n(1)_i + o(n^{-1/2})\right)\mathbf{v}_n'\mathbf{w}_n,
\end{aligned}$$

as $n \to \infty$, a.s.. *Q.E.D.*

PROOF OF THEOREM 4.4: Recall that $\mathcal{A}_n(W) = \sqrt{n}\mathbf{v}_n'(\widehat{\boldsymbol{\Sigma}}_n^s(W) - \boldsymbol{\Sigma}_{n0}^s(W))\mathbf{w}_n$. We have

$$\begin{aligned}
&\mathbf{v}_n'(\widehat{\boldsymbol{\Sigma}}_n^s(\widehat{W}_n^*) - \boldsymbol{\Sigma}_{n0}^s(W_n^*))\mathbf{w}_n \\
&\qquad = \mathbf{v}_n'(\widehat{\boldsymbol{\Sigma}}_n^s(\widehat{W}_n^*) - \boldsymbol{\Sigma}_{n0}^s(\widehat{W}_n^*))\mathbf{w}_n + \mathbf{v}_n'(\boldsymbol{\Sigma}_{n0}^s(\widehat{W}_n^*) - \boldsymbol{\Sigma}_{n0}^s(W_n^*))\mathbf{w}_n \\
&\qquad = n^{-1/2}\mathcal{A}_n(\widehat{W}_n^*) + R_n(\widehat{W}_n^*, W_n^*)
\end{aligned}$$

where again $R_n(\widehat{W}_n^*, W_n^*) = (\widehat{W}_n^* - W_n^*)O(1)$. Further, using

$$\sqrt{n}\mathbf{v}_n'(\widehat{\boldsymbol{\Sigma}}_n^s(\widehat{W}_n^*) - \boldsymbol{\Sigma}_{n0}^s(\widehat{W}_n^*))\mathbf{w}_n = \mathcal{B}_n(\widehat{W}_n^*) + o(1),$$

a.s., we arrive at

$$\begin{aligned}
&\mathbf{v}_n'(\widehat{\boldsymbol{\Sigma}}_n^s(\widehat{W}_n^*) - \boldsymbol{\Sigma}_{n0}^s(W_n^*))\mathbf{w}_n \\
&\qquad = n^{-1/2}\mathcal{A}_n(\widehat{W}_n^*) + R_n(\widehat{W}_n^*, W_n^*) \\
&\qquad = n^{-1/2}(\mathcal{B}_n(\widehat{W}_n^*) + o(1)) + O(\widehat{W}_n^* - W_n^*),
\end{aligned}$$

as $n \to \infty$, a.s., which completes the proof. *Q.E.D.*

PROOF OF THEOREM 4.6: Let $\mathcal{A}_n(W)$ be defined as in (4.7). Arguing as in the proof of Theorem 4.1, we obtain

$$\begin{aligned}
|\mathcal{A}_n(W) - \mathcal{B}_n(W)| \le (1-W)&|\sqrt{n}\mathbf{v}_n'(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)\mathbf{w}_n - \alpha_n(\mathbf{v}_n, \mathbf{w}_n)\overline{B}_n(1)_0| \\
&+ W\left|\mathbf{v}_n'\mathbf{w}_n\sqrt{n}(\|\widehat{\boldsymbol{\Sigma}}_n\|_{tr}^* - \|\boldsymbol{\Sigma}_n\|_{tr}^*) - \mathbf{v}_n'\mathbf{w}_n d_n^{-1/2}\sum_{i=1}^{d_n}\overline{B}_n(1)_i\right|
\end{aligned}$$

The first summand is $o(1)$, a.s., by Theorem 2.1. Under assumption (4.12), the second term can be bounded by

$$\begin{aligned}
R_n &= \left|\mathbf{v}_n'\mathbf{w}_n\sqrt{n}(\|\widehat{\boldsymbol{\Sigma}}_n\|_{tr}^* - \|\widehat{\boldsymbol{\Sigma}}_n\|_{tr}^*) - \mathbf{v}_n'\mathbf{w}_n\sum_{i=1}^{d_n}\overline{B}_n(1)_i\right| \\
&= W|\mathbf{v}_n'\mathbf{w}_n|\left|\sqrt{n}(\|\widehat{\boldsymbol{\Sigma}}_n\|_{tr}^* - \|\widehat{\boldsymbol{\Sigma}}_n\|_{tr}^*) - d_n^{-1/2}\sum_{i=1}^{d_n}\overline{B}_n(1)_i\right| \\
&= o(1),
\end{aligned}$$

as $n \to \infty$, a.s., which completes the proof. *Q.E.D.*

APPENDIX A: NOTATION AND FORMULAS

We denote

$$\sigma_k^2 = E(\epsilon_k^2), \gamma_k = E(\epsilon_k^4).$$

The approximating martingales used to obtain the strong approximations require to control the following quantities. For the reader's convenience, we reproduce them here from [17] as well as some related formulas and results. Let

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

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

and

(A.3) $$\widetilde{f}_{l,i}^{(n)} = \widetilde{f}_{l,i}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = \sum_{j=i}^{\infty} f_{l,j}^{(n)} = \sum_{j=i}^{\infty} \sum_{\nu,\mu=1}^{d_n} v_\nu w_\mu [c_j^{(\nu)} c_{j+l}^{(\mu)} + c_j^{(\mu)} c_{j+l}^{(\nu)}], \qquad l, i = 0, 1, \ldots.$$

LEMMA AND DEFINITION A.1 *Suppose that $\mathbf{v}_n, \mathbf{w}_n$ have uniformly bounded ℓ_1-norm in the sense of equation (1.5). Then Assumption (A) implies*

(A.4) $$\sup_{n\in\mathbb{N}} \sum_{i=1}^{\infty} \sum_{l=0}^{\infty} (\widetilde{f}_{l,i}^{(n)} - \widetilde{f}_{l,i+n'}^{(n)})^2 \le C(n')^{1-\theta}, \qquad \textit{for all } n' = 1, 2, \ldots,$$

(A.5) $$\sup_{n\in\mathbb{N}} \sum_{k=1}^{n'} \sum_{r=0}^{\infty} (\widetilde{f}_{r+k,0}^{(n)})^2 \le C(n')^{1-\theta}, \qquad \textit{for all } n' = 1, 2, \ldots,$$

(A.6) $$\sup_{n\in\mathbb{N}} \sum_{k=1}^{n'} \sum_{l=0}^{\infty} (\widetilde{f}_{l,k}^{(n)})^2 \le C(n')^{1-\theta}, \qquad \textit{for all } n' = 1, 2, \ldots,$$

and there exist

(A.7) $$\alpha_n^2 = \alpha_n^2(\mathbf{v}_n, \mathbf{w}_n) \ge 0, \qquad n \ge 1,$$

such that

(A.8) $$(\widetilde{f}_{00}^{(n)})^2 \sum_{j=1}^{n'} (\gamma_{m'+j} - \sigma_{m'+j}^4) + \sum_{j=1}^{n'} \sum_{l=1}^{j-1} (\widetilde{f}_{j-l,0}^{(n)})^2 \sigma_{m'+j}^2 \sigma_{m'+l}^2 - n'\alpha_n^2 \le C(n')^{1-\theta},$$

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

(A.9) $$\beta_n^2 = \beta_n^2(\mathbf{v}_n, \mathbf{w}_n, \widetilde{\mathbf{v}}_n, \widetilde{\mathbf{w}}_n), \qquad n \ge 1,$$

with

(A.10) $$\widetilde{f}_{0,0}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) \widetilde{f}_{0,0}^{(n)}(\widetilde{\mathbf{v}}_n, \widetilde{\mathbf{w}}_n) \sum_{j=1}^{n'} (\gamma_{m'+j} - \sigma_{m'+j}^4) + \sum_{j=1}^{n'} \sum_{l=1}^{j-1} \widetilde{f}_{j-l,0}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) \widetilde{f}_{j-l,0}^{(n)}(\widetilde{\mathbf{v}}_n, \widetilde{\mathbf{w}}_n) \sigma_{m'+j}^2 \sigma_{m'+l}^2 - n'\beta_n^2(\mathbf{v}_n, \mathbf{w}_n, \widetilde{\mathbf{v}}_n, \widetilde{\mathbf{w}}_n) \overset{n',m'}{<<} (n')^{1-\theta}.$$

ACKNOWLEDGMENTS

Part of this work has been supported by a grant of the first author from Deutsche Forschungsgemeinschaft (DFG), grant No. STE 1034/11-1, which he gratefully acknowledges. 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).

REFERENCES

[1] Andrew R. Barron, Albert Cohen, Wolfgang Dahmen, and Ronald A. DeVore. Approximation and learning by greedy algorithms. *Ann. Statist.*, 36(1):64–94, 2008.

[2] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[3] D. Bosq. *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. Theory and applications.

[4] 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.

[5] Herold Dehling and Walter Philipp. Almost sure invariance principles for weakly dependent vector-valued random variables. *Ann. Probab.*, 10(3):689–701, 1982.

[6] Jianqing Fan, Yingying Fan, and Jinchi Lv. High dimensional covariance matrix estimation using a factor model. *J. Econometrics*, 147(1):186–197, 2008.

[7] Mark Fiecas, Jürgen Franke, Rainer von Sachs, and Joseph Tadjuidje. Shrinkage estimation for multivariate hidden Markov models. *J. Amer. Statist. Assoc.*, to appear, 2016.

[8] Moritz Jirak. Change-point analysis in increasing dimension. *J. Multivariate Anal.*, 111:136–159, 2012.

[9] Tõnu Kollo and Heinz Neudecker. Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices. *J. Multivariate Anal.*, 47(2):283–300, 1993.

[10] Tõnu Kollo and Heinz Neudecker. Corrigendum: "Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices" [J. Multivariate Anal. **47** (1993), no. 2, 283–300; MR1247379 (94j:62040)]. *J. Multivariate Anal.*, 51(1):210, 1994.

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

[12] 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.

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

[14] Weidong Liu and Zhengyan Lin. Strong approximation for a class of stationary processes. *Stochastic Process. Appl.*, 119(1):249–280, 2009.

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

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

[17] A. Steland and R. von Sachs. Large-sample approximations for variance-covariance matrices of high-dimensional time series. *Bernoulli*, in press, 2016.

[18] Terence Tao. A cheap version of the Kabatjanski-Levenstein bound for almost orthogonal vectors.

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

[20] 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.

[21] Wei Biao Wu. Strong invariance principles for dependent random variables. *Ann. Probab.*, 35(6):2294–2320, 2007.

[22] Wei Biao Wu, Yinxiao Huang, and Wei Zheng. Covariances estimation for long-memory processes. *Adv. in Appl. Probab.*, 42(1):137–157, 2010.

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