Journal of Multivariate Analysis 177 (2020) 104582

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# Testing and estimating change-points in the covariance matrix of a high-dimensional time series

Ansgar Steland

*RWTH Aachen University, Institute of Statistics 52065 Aachen, Germany*

A R T I C L E   I N F O

*Article history:*
Received 27 September 2018
Received in revised form 22 November 2019
Accepted 10 December 2019
Available online 17 January 2020

*AMS 2010 subject classifications:*
62E20
62M10
62H99

*Keywords:*
Change-point
CUSUM transform
Data science
High-dimensional statistics
Projection
Spatial statistics
Spiked covariance
Strong approximation
VARMA processes

A B S T R A C T

This paper studies methods for testing and estimating change-points in the covariance structure of a high-dimensional linear time series. The assumed framework allows for a large class of multivariate linear processes (including vector autoregressive moving average (VARMA) models) of growing dimension and spiked covariance models. The approach uses bilinear forms of the centered or non-centered sample variance–covariance matrix. Change-point testing and estimation are based on maximally selected weighted cumulated sum (CUSUM) statistics. Large sample approximations under a change-point regime are provided including a multivariate CUSUM transform of increasing dimension. For the unknown asymptotic variance and covariance parameters associated to (pairs of) CUSUM statistics we propose consistent estimators. Based on weak laws of large numbers for their sequential versions, we also consider stopped sample estimation where observations until the estimated change-point are used. Finite sample properties of the procedures are investigated by simulations and their application is illustrated by analyzing a real data set from environmetrics.



## 1. Introduction

We study methods for the detection of a change-point in a high-dimensional covariance matrix and estimation of its location based on a time series. The proposed procedures investigate estimated bilinear forms of the covariance matrix, in order to test for the presence of a change-point as well as to estimate its location. The bilinear forms use weighting vectors with finite $\ell_1$- resp. $\ell_2$-norms which may even grow slowly as the sample size increases. This approach is natural from a mathematical point of view and has many applications in diverse areas: Analysis of projections onto subspaces spanned by (sparse) principal directions, inferring the dependence structure of high-dimensional sensor data, e.g., from environmental monitoring, testing for a change of the autocovariance function of a univariate series or financial portfolio analysis, to mention a few. These problems have in common that the dimension $d$ can be large and may be even larger than the sample size $n$. The results of this paper allow for this case and do not impose a condition on the growth of the dimension. Multivariate versions of CUSUM statistics are also considered.

The problem to detect changes in a sequence of covariance matrices has been studied by several authors and recently gained increasing interest, although the literature is still somewhat sparse. Going beyond the binary segmentation

*E-mail address:* steland@stochastik.rwth-aachen.de.
*URL:* http://www.isw.rwth-aachen.de.

https://doi.org/10.1016/j.jmva.2019.104582

approach, [7] propose a sparsified segmentation procedure where coordinate-wise CUSUM statistics are thresholded to segment the second-order structure. But these results do not cover significance testing. To test for a covariance change in a time series, [13], who also give some historical references, consider CUSUM and likelihood ratio statistics for fixed dimension $d$ assuming a parametric linear process with Gaussian errors. Their CUSUM statistics, however, require knowledge of the covariance matrix of the innovations when no change is present. [3] studied unweighted and weighted CUSUM change-point tests for a linear process to detect a change in the autocovariance function, but only for a fixed lag. Further, their theoretical results are restricted to the null hypothesis of no change. Kernel methods for this problem have been studied by [34] and [23]. [1] studied break detection in vector time series for fixed dimension and provide an approximation of the limiting distribution of their test statistic, an unweighted CUSUM, if $d$ is large. Contrary, the approach studied in this paper allows for growing dimension $d$ without any constraint such as $d/n \to y \in (0, 1)$, as typically imposed in random matrix theory, $d = O(h(n))$ for some increasing function $h$, e.g., exponential growth as in [2] (which is, however, constrained to i.i.d. samples), or (again for i.i.d. samples) asymptotics for the eigenstructure under the assumption $d/(n\lambda_j) = O(1)$ for the spiked eigenvalues $\lambda_j$, [39], which allows for $d/n \to \infty$ provided the eigenvalues diverge.

It is shown that, for the imposed high-dimensional time series model, (weighted) CUSUM statistics associated to the sample covariance matrix can be approximated by (weighted) Gaussian bridge processes. Under the null hypothesis this follows from [35] and one can also consider an increasing number of such statistics by virtue of the results in [36]. Extensions to several samples (one-factor design) with special emphasis on high-dimensional sensor data are studied in [26]. The asymptotics under a change-point regime, however, is more involved and is provided in this paper. Both single CUSUM statistics and multivariate CUSUM transforms corresponding to a set of projection vectors are studied. The dimension of the time series as well as the dimension of the multivariate CUSUM transform is allowed to grow with the sample size in an unconstrained way. The results of this paper extend [35,36], especially by studying weighted CUSUMs, providing refined martingale approximations and relaxing the conditions on the projection vectors.

Further, consistent estimation of the unknown variance and covariance parameters is studied without the need to estimate eigenstructures. As well known, this essentially would require conditions under which the covariance matrix can be estimated consistently in the Frobenius norm, which needs the restrictive condition $d = o(n)$ on the dimension according to the results of [22] and [30], or requires to assume appropriately constrained models. Estimators for the asymptotic variance and covariance parameters associated to a single resp. a set of CUSUM statistics have already been studied under the no-change hypothesis in [35] and [36]. These estimators are now studied under a change-point model, generalized to deal with two pairs of projection vectors describing the asymptotic covariance between pairs of (weighted) CUSUMs and studied from a sequential viewpoint which allows us to propose stopped-sample estimators using the given sample until the estimated change point. This is achieved by proving a uniform law of large numbers for the sequential estimators.

Closely related to the problem of testing for a change-point is the task of estimating its location. It is shown hat the change-point estimator naturally associated to the weighted or unweighted CUSUM statistic is consistent. As a consequence, the well known iterative binary segmentation algorithm, dating back to [38], can be used to locate multiple change points.

The organization of the paper is as follows. Section 2 introduces the framework, discusses several models appearing as special cases, introduces the proposed methods and discusses how to select the projection vectors. The asymptotic results are provided in Section 3. They cover strong and weak approximations for the (weighted) partial sums of the bilinear forms and for associated CUSUMs as well as consistency theorems for the proposed estimators of unknowns. Section 4 considers the problem to estimate the change-point. Simulations are presented in Section 5. In Section 6 the methods are illustrated by analyzing the dependence structure of ozone measurements from 444 monitors across the United States over a five-year-period. Main proofs are given in Section 7, whereas additional material is deferred to a supplement.

## 2. Model, assumptions and procedures

### 2.1. Notation

Throughout the paper $a_{nk} \overset{n,k}{\ll} b_{nk}$ for two arrays of real numbers means that there exists a constant $C < \infty$, such that $a_{nk} \le C b_{nk}$ for all $n, k$. $(\Omega, \mathcal{A}, \Pr)$ denotes the underlying probability space on which the vector time series is defined. For a logical expression $E$ we let $\mathbf{1}(E)$ denote the associated indicator function. If $A$ is a set, then $\mathbf{1}_A$ is the usual characteristic function, whereas $\mathbf{1}_n$ for $n \in \mathbb{N}$ denotes the $n$-vector with entries 1 and $\mathbf{0}_n$ is the null $n$-vector. $\| \cdot \|_2$ is the vector-2 norm, $\| \cdot \|_{\ell_p}$, $p \in \mathbb{N}$, the $\ell_p$-norm for sequences and $\| \cdot \|_\infty$ the maximum norm for sequences or vectors. $\| \cdot \|_{op}$ denotes the semi norm $\|T\|_{op} = \sup_{f:\|f\|=1} |(f, Tf)\ \|$ for a linear operator $T$ on a Hilbert space with inner product $(\cdot, \cdot)$. $X_n \Rightarrow X$ denotes weak convergence of a sequence of càdlàg processes in the Skorohod space $D[0, 1]$ equipped with the usual metric.

## 2.2. Time series model and assumptions

Let us assume that the coordinates of the vector time series $\boldsymbol{Y}_{ni} = (Y_{ni}^{(1)}, \ldots, Y_{ni}^{(d_n)})^\top$ are given by

$$Y_{ni}^{(\nu)} = Y_{ni}^{(\nu)}(\mathfrak{a}) = \sum_{j=0}^{\infty} a_{nj}^{(\nu)} \epsilon_{n,i-j}, \qquad i \in \{1, \ldots, n\}, \nu \in \{1, \ldots, d_n\}, n \geq 1, \tag{1}$$

for coefficients $\mathfrak{a} = \{a_{nj}^{(\nu)} : j \geq 0, n \geq 1\}$ and independent zero mean errors $\{\epsilon_{ni} : i \in \mathbb{Z}, n \geq 1\}$ satisfying the following two assumptions.

**Assumption (D).** An array $\mathfrak{a} = \{a_{nj}^{(\nu)} : j \geq 0, \nu \in \{1, \ldots, d_n\}, n \geq 1\}$ of real numbers satisfies the decay condition (D), if for some $\theta \in (0, 1/2)$

$$\sup_{n \geq 1} \max_{1 \leq \nu \leq d_n} |a_{nj}^{(\nu)}| \overset{j}{\ll} \min(1, j)^{-3/4-\theta/2}. \tag{2}$$

**Assumption (E).** $\{\epsilon_{nk} : k \in \mathbb{Z}, n \in \mathbb{N}\}$, is an array of independent mean zero random variables with $\sup_{n \geq 1} \sup_{k \in \mathbb{Z}} \mathrm{E} |\epsilon_{nk}|^{4+\delta} < \infty$ and moment arrays $\sigma_{nk}^2 = \mathrm{E}(\epsilon_{nk}^2)$, $\gamma_{nk} = \mathrm{E}(\epsilon_{nk}^3)$, $1 \leq k \leq n$, $n \geq 1$, satisfying

$$\frac{1}{\ell} \sum_{i=1}^{\ell} i |\sigma_{ni}^2 - s_{n1}^2| = O(\ell^{-\beta}), \qquad \frac{1}{\ell} \sum_{i=1}^{\ell} i |\gamma_{ni} - \gamma_n| = O(\ell^{-\beta}),$$

for some $\beta > 1 + \theta$ and sequences $\{s_{n1}\}$ and $\{\gamma_n\}$.

The assumptions on $\sigma_{ni}^2$ and $\gamma_{ni}$ allow for a certain degree of inhomogeneity of the second and third moments. Especially, under the change-point model described below, where the coefficients of the linear processes change after the change-point $\tau = \tau_n$, these assumptions cover weak effects of the change on the second resp. third moments. An example satisfying the conditions is given by

$$\sigma_{ni}^2 = s_{n1}^2 + \frac{\kappa_i}{i} \Delta_{\sigma^2, ni}, \quad \Delta_{\sigma^2, ni} = \mathbf{1}(i \leq \tau)\sigma_0^2 + \mathbf{1}(i > \tau)\sigma_1^2,$$

for two positive constants $\sigma_0^2 \neq \sigma_1^2$ and $\kappa_i \in \mathbb{R}$, $i \geq 1$, with $\kappa_i = o(i)$ and $\sum_{i=1}^{\ell} |\kappa_i| \sim l^{1+\beta}$.

## 2.3. Spiked covariance model

The spiked covariance model is a common framework to study estimation of the eigenstructure for high-dimensional data. For $r \in \mathbb{N}$ let $\lambda_1 > \cdots > \lambda_r > 0$ and let $\boldsymbol{u}_{nj} = (u_{nj}^{(\nu)})_{\nu=1}^{d} \in \mathbb{R}^d$, $j \in \{1, \ldots, r\}$, be orthonormal vectors with $\|\boldsymbol{u}_{nj}\|_{\ell_1} \leq C$ for $j \in \{1, \ldots, r\}$. Assume that

$$\boldsymbol{\Sigma}_n = \sum_{j=1}^{r} \lambda_j \boldsymbol{u}_{nj} \boldsymbol{u}_{nj}^\top + \sigma^2 I_{d_n}. \tag{3}$$

The $r$ leading eigenvalues of $\boldsymbol{\Sigma}_n$ under model (3) are $\lambda_j + \sigma^2$, $j \in \{1, \ldots, r\}$, and represent spikes in the spectrum, which is otherwise flat and given by $\sigma^2$. The assumption that the eigenvectors are $\ell_1$-bounded is common in high-dimensional statistics, especially when assuming a spiked covariance model: [17] have shown that principal component analysis (PCA) generates inconsistent estimates of the leading eigenvectors if $d/n \to y \in (0, 1)$, which motivated developments on sparse PCA. Minimax bounds for sparse PCA have been studied by [4] under $\ell_q$-constraints on the eigenvectors for $0 < q < 2$. For example, the simple diagonal thresholding estimator $\widehat{\boldsymbol{u}}_{nj}^{th}$ of the $j$th leading eigenvector $\boldsymbol{u}_{nj}$ of [17] satisfies $\mathrm{E}\|\widehat{\boldsymbol{u}}_{nj}^{th} - \widehat{\boldsymbol{u}}_{nj}^{th\top} \boldsymbol{u}_{nj} \boldsymbol{u}_{nj}\|_2^2 = O(n^{-1/4})$ and the iterated version of [24] attains the optimal rate $O(n^{(1-q/2)})$, see also [28]. An $\ell_1$ sparseness assumption on the eigenvectors is weaker than the (joint) $k$-sparseness condition on the row support of matrix of eigenvectors imposed in [6], who study optimal estimation under the spectral norm.

Model (3) can be described in terms of (1): Let $c_{n,r+\nu}^{(\nu)} = \sigma^2$, $c_{n,j-1}^{(\nu)} = \lambda_j^{1/2} u_{nj}^{(\nu)}$, $j \in \{1, \ldots, r\}$, and $c_{nj}^{(\nu)} = 0$, otherwise. Then for $\epsilon_t$ i.i.d $\mathcal{N}(0, 1)$ the MA$(r+d-1)$ series $Y_{nt}^{(\nu)} = \sum_{j=0}^{r-1} c_{nj}^{(\nu)} \epsilon_{t-j} + \sigma \epsilon_{t-r-\nu}$, $\nu \in \{1, \ldots, d\}$, have the covariance matrix (3). The decay condition (D) follows from $\sup_{1 \leq n} \max_{1 \leq \nu \leq r} |u_{nj}^{(\nu)}| \leq \sup_{n \geq 1} \|\boldsymbol{u}_{nj}\|_{\ell_1}$.

We may conclude that our methodology covers the above spiked covariance under which sparse PCA provides consistent estimates of the leading eigenvectors, which are an attractive choice for the projection vectors on which the proposed change-point procedures are based on. The literature on such consistency results is, however, not yet matured and typically assumes i.i.d. data vectors, whereas the framework studied here considers time series.

*2.4. Multivariate linear time series and VARMA processes*

The above linear process framework is general enough to host classes of multivariate linear processes and vector autoregressive models with respect to a $q$-variate noise process, $q \in \mathbb{N}$. These processes are usually studied for a sequence of innovations, but since our constructions work for arrays, we consider this setting.

*Multivariate linear processes:* Let $0 = r_1 \le r_2 \le \cdots \le r_q$ be integers and define the $q$-variate innovations

$$\boldsymbol{\epsilon}_{ni} = (\epsilon_{n,i-r_1}, \ldots, \epsilon_{n,i-r_q})^\top, \qquad i \ge 1, n \ge 1,$$

based on $\{\epsilon_{ni} : i \ge 1, n \ge 1\}$. If $\epsilon_{ni}$ have homogeneous variances, then $E(\boldsymbol{\epsilon}_{n0}\boldsymbol{\epsilon}_{nk}^\top) \ne \mathbf{0}$ iff. $k \in \{r_j - r_i : 1 \le i, j \le q\}$, $k \ne 0$, such that for large enough $r_j, j \ge 2$, the innovations are arbitrarily close to white noise. Let $\boldsymbol{B}_{nj} = (\boldsymbol{b}_{nj,1}, \ldots, \boldsymbol{b}_{nj,d_n})^\top$, be $(d_n \times q)$-dimensional matrices with row vectors $\boldsymbol{b}_{nj,\nu} = (b_{nj}^{(\nu,1)}, \ldots, b_{nj}^{(\nu,q)})^\top$, $\nu \in \{1, \ldots, d_n\}$, for $j \ge 0$. Then the $d_n$-dimensional linear process

$$\boldsymbol{Z}_{ni} = \sum_{j=0}^{\infty} \boldsymbol{B}_{nj}\boldsymbol{\epsilon}_{n,i-j}$$

has coordinates $Z_{ni}^{(\nu)} = \sum_{j=0}^{\infty}\sum_{\ell=1}^{q} b_{nj}^{(\nu,\ell)}\epsilon_{n,i-r_\ell-j}$, which attain the representation

$$Z_{ni}^{(\nu)} = \sum_{k=0}^{\infty}\left(\sum_{\ell=1}^{q} \mathbf{1}(k \ge r_\ell) b_{n,k-r_\ell}^{(\nu,\ell)}\right)\epsilon_{n,i-k}, \tag{4}$$

$\nu \in \{1, \ldots, d_n\}$. If we assume that the elements $b_{nj}^{(\nu,\ell)}$ of the coefficient matrices $\boldsymbol{B}_{nj}$ satisfy the decay condition

$$\max_{\nu \ge 1}\left|b_{nj}^{(\nu,\ell)}\right| \overset{n}{\ll} (j + r_\ell)^{-3/4-\theta/2},$$

then the coefficients $c_{nk}^{(Z,\nu)} = \sum_{\ell=1}^{q}\mathbf{1}(k \ge r_\ell) b_{n,k-r_\ell}^{(\nu,\ell)}$ of the series (4) satisfy $\sup_{n\ge1}\max_{\nu\ge1}|c_{nk}^{(Z,\nu)}| \ll k^{-3/4-\theta/2}$, i.e., Assumption (D) holds. In this construction the lags $r_1, \ldots, r_q$ used to define the $q$-variate innovation process may depend on $n$.

We may go beyond the above near white noise $q$-variate innovations and consider $d_n$-dimensional linear processes with mean zero innovations $\boldsymbol{e}_{ni}$, $i \ge 1$, with a covariance matrix close to some $\boldsymbol{V} > 0$: Let

$$\boldsymbol{Z}_{ni} = \sum_{j=0}^{\infty} \boldsymbol{B}_{nj}\boldsymbol{P}\boldsymbol{e}_{n,i-j}, \qquad \boldsymbol{e}_{ni} = \boldsymbol{V}^{1/2}\boldsymbol{\epsilon}_{ni}, \qquad i \ge 1, n \in \mathbb{N}, \tag{5}$$

where

$$\boldsymbol{\epsilon}_{ni} = (\epsilon_{n,i-r_1}, \ldots, \epsilon_{n,i-r_{d_n}})^\top, \qquad i \ge 1, n \in \mathbb{N}, \tag{6}$$

for $0 = r_1 < \cdots < r_{d_n}$, $\boldsymbol{P}$ is a full rank $q \times d_n$ matrix and $\boldsymbol{B}_{nj}$ are $d_n \times q$ coefficient matrices as above, i.e., with elements satisfying the decay condition. $\boldsymbol{P}$ is used to reduce the dimensionality. Let $\boldsymbol{P}\boldsymbol{V}^{1/2} = \sum_{i=1}^{q}\pi_{ni}\boldsymbol{l}_{ni}\boldsymbol{r}_{ni}^\top$ be the singular value decomposition of $\boldsymbol{P}\boldsymbol{V}^{1/2}$ with singular values $\pi_{ni}$, left singular vectors $\boldsymbol{l}_{ni} \in \mathbb{R}^q$ and right singular vectors $\boldsymbol{r}_{ni} = (r_{ni1}, \ldots, r_{nid_n})^\top \in \mathbb{R}^d$ satisfying $\|\boldsymbol{l}_{ni}\|_{\ell_2} = \|\boldsymbol{r}_{ni}\|_{\ell_2} = 1$, $i \in \{1, \ldots, d_n\}$, $n \ge 1$. Then $\boldsymbol{B}_{nj}\boldsymbol{P}\boldsymbol{V}^{1/2} = \sum_{i=1}^{q}\pi_{ni}\boldsymbol{B}_{nj}\boldsymbol{l}_{ni}\boldsymbol{r}_{ni}^\top$, and the element at position $(\nu, \ell)$ of the latter matrix is given by $\sum_{i=1}^{q}\pi_{ni}\boldsymbol{b}_{nj,\nu}^\top\boldsymbol{l}_{ni}r_{ni\ell}$ which is $\overset{\nu}{\ll} j^{-3/4-\theta/2}$ if the eigenvalues and eigenvectors are bounded. Therefore, the class of processes (5) is a special case of (1).

The case $q = q_n \to \infty$, especially $q = d_n$ leading to the usual definition of a $d_n$-dimensional linear process, can be allowed for when imposing the conditions

$$\max_{\nu,\mu\ge1}\left|b_{nj}^{(\nu,\mu)}\right| \overset{n}{\ll} (j + 2r_\ell)^{-3/2-\varpi-(\theta+\varpi)}(\nu\mu)^{\varpi(-1/2-\theta)} \qquad \text{and} \qquad \sup_{n\ge1}\sum_{\ell=1}^{\infty} r_\ell^{-3/4-\theta/2} < \infty \tag{7}$$

with $\varpi = 0$ and assuming that the operators $\boldsymbol{B}_{nj}\boldsymbol{P}\boldsymbol{V}^{1/2}$, $n \ge 1$, are trace class operators in the sense that $\sum_i|\pi_{ni}| = O(1)$, with eigenvectors satisfying $\|\boldsymbol{l}_{ni}\|_{\ell_1}, \|\boldsymbol{r}_{ni}\|_{\ell_1} \overset{n,i}{\ll} 1$. For $q = d_n$ we let $\boldsymbol{P} = I$ such that $\boldsymbol{l}_{ni} = \boldsymbol{r}_{ni}$ are the eigenvectors and $\pi_{ni}$ the eigenvalues of $\boldsymbol{V}$. Then $\sup_{n\ge1}\max_{\nu\ge1}\left|c_{nj}^{(Z,\nu)}\right| \ll j^{-3/4-\theta/2}\sum_{\ell=1}^{\infty} r_\ell^{-3/4-\theta/2} \ll j^{-3/4-\theta/2}$ verifying (D), as shown in the supplement. The $\ell_1$ constraint on the eigenvectors can be omitted when imposing the stronger condition $\sup_{\nu\ge1}\sum_{j=1}^{d_n}|b_{n,k-r_\ell}^{(\nu,j)}| \overset{n}{\ll} (k + r_\ell)^{-3-\theta}$ on the coefficient matrices. For details see the supplement.

*VARMA Models:* Let us consider a $d_n$-dimensional zero mean VARMA$(p, r)$ process

$$\boldsymbol{Y}_{ni} = \boldsymbol{A}_{n1}\boldsymbol{Y}_{n,i-1} + \cdots + \boldsymbol{A}_{np}\boldsymbol{Y}_{n,i-p} + \boldsymbol{M}_{n1}\boldsymbol{\epsilon}_{n,i-1} + \cdots + \boldsymbol{M}_{nr}\boldsymbol{\epsilon}_{n,i-r} + \boldsymbol{\epsilon}_{ni},$$

with colored $d_n$-variate innovations as in (5). $\boldsymbol{A}_{n1}, \ldots, \boldsymbol{A}_{np}$ and $\boldsymbol{M}_{n1}, \ldots, \boldsymbol{M}_{nr}$ are $(d_n \times d_n)$ coefficient matrices. Let us assume that each of these coefficient matrices satisfies (7) with $\varpi = 1$ for some $\theta > 0$, when denoting its elements by

$b_{nj}^{(\nu,\ell)}$, $1 \le \nu, \ell \le d_n$. Recall that the process is stable, if $\det(I_{d_n} - \boldsymbol{A}_{n1} z - \cdots - \boldsymbol{A}_{np} z^p) \neq 0$ for $|z| \le 1$. Then the operator $\boldsymbol{A}(L) = I_{d_n} - \sum_{j=1}^{p} \boldsymbol{A}_{nj} L^j$, where $L$ denotes the lag operator, is invertible, the coefficient matrices, $\boldsymbol{D}_{nj}$, of $\boldsymbol{\Psi}(L) = \boldsymbol{A}(L)^{-1}$ are absolutely summable, and one obtains the MA representation $\boldsymbol{Y}_{ni} = \left(\sum_{j=0}^{\infty} \boldsymbol{D}_{nj} L^j\right) \left(\sum_{k=1}^{r} \boldsymbol{M}_{nk} L^k\right) \boldsymbol{\epsilon}_{n,i-j} = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_{nj} \boldsymbol{\epsilon}_{n,i-j}$. As well known, the coefficient matrices, $\boldsymbol{\Phi}_{nj}$, can be calculated using the recursion $\boldsymbol{\Phi}_{n0} = I_{d_n}$, $\boldsymbol{\Phi}_{nj} = \boldsymbol{M}_{nj} + \sum_{k=1}^{j} \boldsymbol{A}_{nk} \boldsymbol{\Phi}_{n,j-k}$, $j \ge 1$, where $\boldsymbol{M}_{nj} = \boldsymbol{0}$ for $j > q$. Using these formulas one can show that the coefficient matrices, $\boldsymbol{\Phi}_{nj}$, of the MA representation satisfy (7) when denoting its elements by $b_{nj}^{(\nu,\ell)}$, and therefore the VARMA coordinate processes $Y_{ni}^{(\nu)}$, $1 \le \nu \le d_n$, with innovations (6) satisfy the decay condition (D), i.e., if the elements of the coefficient matrices are of the order $(j + 2r_\ell)^{-5/2-2\theta} (\nu\mu)^{-1/2-\theta}$, then the elements of the matrices of the MA representation are of the order $(j + 2r_\ell)^{-3/2-\theta}$.

Another interesting class of time series to be studied in future work are factor models, which are of substantial interest in econometrics. For detection of changes resp. breaks we refer to [5,14] and [15], amongst others.

*2.5. Change-point model and procedures*

The change-point model studied in this paper considers a change of the coefficients defining the linear processes. Nevertheless, all procedures neither require their knowledge nor their estimation. So let $\mathfrak{b} = \{b_{nj}^{(\nu)} : j \ge 0, \nu \in \{1, \dots, d_n\}, n \ge 1\}$ and $\mathfrak{c} = \{c_{nj}^{(\nu)} : j \ge 0, \nu \in \{1, \dots, d_n\}, n \ge 1\}$ be two different coefficient arrays satisfying the decay assumption and put

$$\boldsymbol{\Sigma}_{n0} = \boldsymbol{\Sigma}_n(\mathfrak{b}) = \operatorname{Var}(\boldsymbol{Y}_n(\mathfrak{b})), \qquad \boldsymbol{\Sigma}_{n1} = \boldsymbol{\Sigma}_n(\mathfrak{c}) = \operatorname{Var}(\boldsymbol{Y}_n(\mathfrak{c})).$$

It is further assumed that $\mathfrak{b}$ and $\mathfrak{c}$ are such that

$$\mathfrak{b} \neq \mathfrak{c} \Rightarrow \boldsymbol{\Sigma}_{n0} \neq \boldsymbol{\Sigma}_{n1}, \qquad n \in \mathbb{N}. \tag{8}$$

We will study CUSUM type procedures based on quadratic and bilinear forms of sample analogs of those variance–covariance matrices, in order to detect a change from $\boldsymbol{\Sigma}_{n0}$ to $\boldsymbol{\Sigma}_{n1}$. Let $\mathcal{V}_n = \{(\boldsymbol{x}_n, \boldsymbol{y}_n) \in \mathbb{R}^{d_n} \times \mathbb{R}^{d_n} : \boldsymbol{x}_n^\top \boldsymbol{\Sigma}_{n0} \boldsymbol{y}_n \neq \boldsymbol{x}_n^\top \boldsymbol{\Sigma}_{n1} \boldsymbol{y}_n\}$. Assumption (8) ensures that $\mathcal{V}_n \neq \emptyset$.

The change-point model for the high-dimensional time series is now as follows. For some change-point $\tau \in \{1, \dots, n\}$ it holds

$$\boldsymbol{Y}_{ni} = \boldsymbol{Y}_{ni}(\mathfrak{b}) \mathbf{1}(i \le \tau) + \boldsymbol{Y}_{ni}(\mathfrak{c}) \mathbf{1}(i > \tau), \qquad 1 \le i \le n, \tag{9}$$

with underlying error terms $\epsilon_{ni}$, $1 \le i \le n$, $n \ge 1$, satisfying Assumption (E). Our results on estimation of $\tau$, however, assume that the change occurs after a certain fraction of the sample by requiring that

$$\tau = \lfloor n\vartheta \rfloor, \tag{10}$$

for some $\vartheta \in (0, 1)$. We are interested in testing the change-point problem

$$H_0 : \tau = n \qquad \text{versus} \qquad H_1 : \tau < n,$$

which implies a change in the second moment structure of the vector time series $\boldsymbol{Y}_{n1}, \dots, \boldsymbol{Y}_{nn}$ when $H_1$ holds, and in estimation of the change-point to locate the change. Under the null hypothesis the covariance matrix of $\boldsymbol{Y}_{ni}$, $1 \le i \le n$, is given by $\operatorname{Var}(\boldsymbol{Y}_{n1}(\mathfrak{b})) = \boldsymbol{\Sigma}_{n0}$, whereas it changes under the alternative hypothesis from $\boldsymbol{\Sigma}_{n0}$ to $\operatorname{Var}(\boldsymbol{Y}_{n,\tau+1}(\mathfrak{c})) = \boldsymbol{\Sigma}_{n1}$. If $(\boldsymbol{v}_n, \boldsymbol{w}_n) \in \mathcal{V}_n$, then the change is present in the sequence of the associated quadratic forms, $\sigma_n^2[k] = \boldsymbol{v}_n^\top \operatorname{Var}(\boldsymbol{Y}_{nk}) \boldsymbol{w}_n$, $1 \le k$, which change from $\boldsymbol{v}_n^\top \boldsymbol{\Sigma}_{n0} \boldsymbol{w}_n$ to $\boldsymbol{v}_n^\top \boldsymbol{\Sigma}_{n1} \boldsymbol{w}_n$ if $\tau < n$, and the change-point test below will be based on an estimator of that bilinear form. A natural condition to ensure that this relationship holds asymptotically, yielding consistency of the proposed test, is

$$\inf_{n \ge 1} |\Delta_n| > 0, \qquad \Delta_n := \boldsymbol{v}_n^\top \boldsymbol{\Sigma}_{n0} \boldsymbol{w}_n - \boldsymbol{v}_n^\top \boldsymbol{\Sigma}_{n1} \boldsymbol{w}_n. \tag{11}$$

We shall, however, also discuss in Section 3.2 more general conditions for the detectability of a change.

To introduce the proposed procedures, define the partial sums of the outer products $\boldsymbol{Y}_{ni} \boldsymbol{Y}_{ni}^\top$,

$$\boldsymbol{S}_{nk} = \sum_{i \le k} \boldsymbol{Y}_{ni} \boldsymbol{Y}_{ni}^\top, \qquad k \ge 1,$$

such that $k^{-1} \boldsymbol{S}_{nk}$ is the sample variance–covariance matrix using the data $\boldsymbol{Y}_{n1}, \dots, \boldsymbol{Y}_{nk}$. Let

$$U_{nk} = \boldsymbol{v}_n^\top \boldsymbol{S}_{nk} \boldsymbol{w}_n, \qquad k \ge 1, n \ge 1.$$

Consider the CUSUM-type statistic,

$$C_n = C_n(\boldsymbol{v}_n, \boldsymbol{w}_n) = \max_{1 \le k < n} \frac{1}{\sqrt{n}} \left| \boldsymbol{v}_n^\top \left( \boldsymbol{S}_{nk} - \frac{k}{n} \boldsymbol{S}_{nn} \right) \boldsymbol{w}_n \right|.$$

The large sample approximations for $C_n$ obtained in [35] under $H_0$, and generalized in this paper, imply that $C_n$ can be approximated by a Brownian bridge process, $B^0$. Hence, we can reject the null hypothesis of no change at the asymptotic level $\alpha \in (0,1)$, if

$$T_n > K_{1-\alpha}^{-1}, \qquad T_n = T_n(\boldsymbol{v}_n, \boldsymbol{w}_n) = \max_{1\le k<n} \frac{1}{\widehat{\alpha}_n(\mathfrak{b})\sqrt{n}} \left| \boldsymbol{v}_n^\top \left( \boldsymbol{S}_{nk} - \frac{k}{n}\boldsymbol{S}_{nn} \right) \boldsymbol{w}_n \right|, \tag{12}$$

where $\widehat{\alpha}_n(\mathfrak{b})$ is a consistent estimator for the asymptotic standard deviation $\alpha_n(\mathfrak{b})$ associated to the series $\boldsymbol{v}_n^\top \boldsymbol{S}_{nk}\boldsymbol{w}_n$, and $K_u^{-1}$ is the $u$-quantile, $u \in (0,1)$, of the Kolmogorov distribution function, $K(z) = 1 - \sum_{i=1}^\infty (-1)^{i-1} \exp(-2i^2z^2)$, $z \in \mathbb{R}$. One may also use a weighted CUSUM test

$$C_n(g) = \max_{1\le k<n} \frac{1}{\sqrt{n}g(k/n)} \left| \boldsymbol{v}_n^\top \left( \boldsymbol{S}_{nk} - \frac{k}{n}\boldsymbol{S}_{nn} \right) \boldsymbol{w}_n \right|$$

for some weight function $g$, whose role is to compensate for the fact that the centered cumulated sums get small near the boundaries. The results of Section 3.2 provide large sample approximations for a large class of weighting functions. An attractive choice would be the weight function $g(t) = \sqrt{t(1-t)}$, but the corresponding supremum of the standardized Brownian bridge, $B^0(t)/\sqrt{t(1-t)}$, $0 < t < 1$, is not well defined due to the law of the iterated logarithm (LIL), requiring then to use Gumbel-type extreme value asymptotics, [9], known to converge slowly, see, e.g., [12]. For a discussion of the class of proper weight functions ensuring that $B^0(t)/g(t)$ is a.s. finite we refer to [8]. One may use the weight function $[t(1-t)]^\beta$ for some $0 < \beta < 1/2$ or any weight function $g$ satisfying

$$g(t) \ge C_g[t(1-t)]^\beta, \;\; 0 \le t \le 1, \;\; \text{for some constant } C_g. \tag{13}$$

Therefore, one rejects the no-change null hypothesis, if

$$T_n(g) > q_g(1-\alpha), \tag{14}$$

where $q_g$ denotes the quantile function of the law of $\sup_{0<t<1} |B^0(t)|/g(t)$. As studied in [12], one may also standardize the unweighted CUSUM statistic by its maximizing point, i.e., substitute $g(k/n)$ by $\sqrt{\widehat{\tau}_n(1-\widehat{\tau}_n)}$. The associated Brownian bridge standardized by its argmax attains a density which has been explicitly calculated in [12].

When the assumption that the vector time series has mean zero is in doubt, one may modify the above procedures by taking the cumulated outer products of the centered series, $\widetilde{\boldsymbol{S}}_{nk} = \sum_{i\le k}(\boldsymbol{Y}_{ni} - \overline{\boldsymbol{Y}}_n)(\boldsymbol{Y}_{ni} - \overline{\boldsymbol{Y}}_n)^\top$, where $\overline{\boldsymbol{Y}}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{Y}_{ni}$. The associated weighted CUSUM statistics are then given by

$$\widetilde{C}_n(g) = \max_{1\le k<n} \frac{1}{\sqrt{n}g(k/n)} \left| \boldsymbol{v}_n^\top \left( \widetilde{\boldsymbol{S}}_{nk} - \frac{k}{n}\widetilde{\boldsymbol{S}}_{nn} \right) \boldsymbol{w}_n \right|, \quad \widetilde{T}_n(g) = \frac{\widetilde{C}_n(g)}{\widehat{\alpha}_n(\mathfrak{b})},$$

and the null hypothesis is rejected using the rule (14) with $T_n(g)$ replaced by $\widetilde{T}_n(g)$.

To estimate the unknown change-point $\tau$, we propose to use the estimator

$$\widehat{\tau}_n = \operatorname*{argmax}_{1\le k<n} \frac{1}{g(k/n)n} \left| \boldsymbol{v}_n^\top \left( \boldsymbol{S}_{nk} - \frac{k}{n}\boldsymbol{S}_{nn} \right) \boldsymbol{w}_n \right|.$$

Based on the estimator $\widehat{\tau}_n$ of the change-point, one may also estimate the nuisance parameter $\alpha_n^2(\mathfrak{b})$ by $\widehat{\alpha}^2_{\widehat{\tau}_n}(\mathfrak{b})$.

For $L$ pairs of projection vectors $\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj}$, $j \in \{1,\dots,L\}$, consider the associated CUSUM transform

$$\boldsymbol{C}_n = \left( C_n(\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj}) \right)_{j=1}^L, \qquad \boldsymbol{T}_n = \left( T_n(\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj}) \right)_{j=1}^L.$$

Observe that this transform differs from the transform studied in [40], where the statistics are calculated coordinate-wise and the transform is given by the corresponding $d$ CUSUM trajectories.

We wish to test the null hypothesis of no change w.r.t. to $\{\boldsymbol{v}_n, \boldsymbol{w}_n\}$

$$H_0 : \boldsymbol{v}_{nj}^\top \mathrm{Var}(\boldsymbol{Y}_{n\tau})\boldsymbol{w}_{nj} = \boldsymbol{v}_{nj}^\top \mathrm{Var}(\boldsymbol{Y}_{n,\tau+1})\boldsymbol{w}_{nj}, j \in \{1,\dots,L\},$$

against the alternative hypothesis that, induced by a change at $\tau < n$, at least one bilinear form changes (assuming the projections are appropriately selected),

$$H_1 : \exists j \in \{1,\dots,L\} : \boldsymbol{v}_{nj}^\top \mathrm{Var}(\boldsymbol{Y}_{n\tau})\boldsymbol{w}_{nj} \ne \boldsymbol{v}_{nj}^\top \mathrm{Var}(\boldsymbol{Y}_{n,\tau+1})\boldsymbol{w}_{nj}.$$

As a global (omnibus) test one may reject $H_0$ at the asymptotic significance level $\alpha$, if

$$Q_n = (\boldsymbol{T}_n - \boldsymbol{\mu}_n^*)^\top (\widehat{\boldsymbol{\Sigma}}_n^{\boldsymbol{B}})^- (\boldsymbol{T}_n - \boldsymbol{\mu}_n^*) > q_{vm}(1-\alpha). \tag{15}$$

Here $Q_n$ is a non-standard quadratic form, as it is based on the CUSUMs instead of a multivariate statistic which is asymptotically normal, $\boldsymbol{\mu}_n^* = \left( \max_{1\le k<n} \mathrm{E}\max_{1\le k<n} |\overline{B}^0(k/n)/g(k/n)| \right)_{j=1}^L$, $(\widehat{\boldsymbol{\Sigma}}_n^{\boldsymbol{T}})^-$ is the Moore–Penrose generalized inverse of $\widehat{\boldsymbol{\Sigma}}_n^{\boldsymbol{T}} = \left( \widehat{\beta}_n^2(j,k)\widehat{\beta}_n^{-1}(k,k)\widehat{\beta}_n^{-1}(j,j) \right)_{\substack{1\le j\le L \\ 1\le k\le L}}$ and $q_{mv}(p)$ denotes the $p$-quantile of the simulated distribution of $Q_n$

using a Monte Carlo estimate of $\mathrm{E}\max_{1\le k<n}|\overline{B}^0(k/n)/g(k/n)|$; the estimators $\widehat{\beta}_n^2(j,k)$ of the asymptotic covariance of the $j$th and $k$th coordinate of the CUSUM transform $\boldsymbol{C}_n$ are defined in the next section and are calculated from a learning sample. It is worth mentioning that the statistic $Q_n$ can be used to test for a change in the subspace $\mathrm{span}\{\boldsymbol{v}_{n1},\dots,\boldsymbol{v}_{nL}\}$ by putting $\boldsymbol{w}_{ni}=\boldsymbol{v}_{ni}$, $i\in\{1,\dots,n\}$.

### 2.6. Choice of the projections

The question arises how to choose the projection vectors $\boldsymbol{v}_n,\boldsymbol{w}_n$. Their choice may depend on the application and some of the examples discussed in the supplement illustrate this. In some applications selecting them from a known basis may be the method of choice. In low- and high-dimensional multivariate statistics it is, however, a common statistical tool to project data vectors onto a lower dimensional subspace spanned by (sparse) directions (axes) $\boldsymbol{v}_n^{(1)},\dots,\boldsymbol{v}_n^{(K)}$. These directions can be obtained from a fixed basis or by a (sparse) principal component analysis using a learning sample. The projection is determined by the new coordinates $\boldsymbol{v}_n^{(i)\top}\boldsymbol{Y}_n$, $i\in\{1,\dots,K\}$, for simplicity also called projections, and represent a lower dimensional compressed approximation of $\boldsymbol{Y}_n$. The uncertainty of its coordinates, i.e., of its position in the subspace, can be measured by the variances $\boldsymbol{v}_n^{(i)\top}\boldsymbol{\Sigma}_n\boldsymbol{v}_n^{(i)}$. Clearly, it is of interest to test for the presence of a change-point in the second moment structure of these new coordinates by analyzing the bilinear forms $\boldsymbol{v}_n^{(i)\top}\boldsymbol{\Sigma}_n\boldsymbol{v}_n^{(j)}$, $1\le i,j\le K$. Also observe that one may analyze the spectrum, since for eigenvectors $\boldsymbol{v}_n^{(i)}$ the associated eigenvalue is given by $\boldsymbol{v}_n^{(i)\top}\boldsymbol{\Sigma}_n\boldsymbol{v}_n^{(i)}$.

The question under which conditions PCA or sparse PCA is consistent has been studied by various authors. The classic Davis–Kahan theorem, see [10] and [41] for a statistical version, relates this to consistency of the sample covariance matrix in the Frobenius norm, which generally does not hold under high-dimensional regimes without additional assumptions, and minimal-gap conditions on the eigenvalues. Standard PCA is known to be inconsistent, if $d/n\to y\in(0,\infty]$, where here and in the following discussion a possible dependence of $d$ on $n$ is suppressed. Under certain spiked covariance models consistency can be achieved, see [17] if $d=o(n)$, [27] under the condition $d/n\to\gamma\in(0,1)$ and [19] for $n$ fixed and $d\to\infty$. Sparse principal components, first formally studied by [18] using lasso techniques, are strongly motivated by data-analytic aspects, e.g., by simplifying their interpretation, since linear combinations found by PCA typically involve all variables. Consistency has been studied under different frameworks, usually assuming additional sparsity constraints on the true eigenvectors (to ensure that their support set can be identified) and/or growth conditions on the eigenvalues (to ensure that the leading eigenvalues are dominant in the spectrum). We refer to [32] for simple thresholding sparse PCA when $n$ is held fixed and $d\to\infty$, [4] for results on minimax rates when estimating the leading eigenvectors under $\ell_q$-constraints on the eigenvectors and fixed eigenvalues, whereas [6] provide minimax bounds assuming at most $k$ entries of the eigenvectors are non-vanishing and [39] derives asymptotic distributions allowing for diverging eigenvalues and $d/n\to\infty$.

To avoid that a change is not detectable because it takes place in a subspace of the orthogonal complement of the chosen projection vectors, a simple approach used in various areas is to take random projections. For example, one may draw the projection vectors from a fixed basis or, alternatively, sample them from a distribution such as a Dirichlet distribution or an appropriately transformed Gaussian law. Random projections of such kind are also heavily used in signal processing and especially in compressed sensing, by virtue of the famous distributional version of the Johnson–Lindenstrauss theorem, see [16]. This theorem states that any $n$ points in a Euclidean space can be embedded into $O(\varepsilon^2\log(1/\delta))$ dimensions such that their distances are preserved up to $1\pm\varepsilon$, with probability larger than $1-\delta$. This embedding can be constructed with $\ell_0$-sparsity $O(\varepsilon^{-1}\log(1/\delta))$ of the associated projection matrix, see [20].

This discussion is continued in the next section after Theorem 2 and related to the change-point asymptotics established there.

## 3. Asymptotics

The asymptotic results comprise approximations of the CUSUM statistics and related processes by maxima of Gaussian bridge processes, consistency of Bartlett type estimators of the asymptotic covariance structure of the CUSUMs, stopped sample versions of those estimators and consistency of the proposed change-point estimator.

### 3.1. Preliminaries

To study the asymptotics of the proposed change-point test statistics both under $H_0$ and $H_1$, we consider the two-dimensional partial sums,

$$\boldsymbol{U}_{nk}=\begin{pmatrix}\boldsymbol{U}_{nk}^{(1)}\\ \boldsymbol{U}_{nk}^{(2)}\end{pmatrix}=\sum_{i\le k}\begin{pmatrix}Y_{ni}(\boldsymbol{v}_n^\top\boldsymbol{b}_n)Y_{ni}(\boldsymbol{w}_n^\top\boldsymbol{b}_n)\\ Y_{ni}(\boldsymbol{v}_n^\top\boldsymbol{c}_n)Y_{ni}(\boldsymbol{w}_n^\top\boldsymbol{c}_n)\end{pmatrix}\tag{16}$$

and their centered versions,

$$\boldsymbol{D}_{nk}=\boldsymbol{U}_{nk}-\mathrm{E}(\boldsymbol{U}_{nk}),\tag{17}$$

for $k, n \geq 0$, where for brevity $Y_{ni}(\boldsymbol{v}_n^\top \boldsymbol{b}_n)$ and $Y_{ni}(\boldsymbol{v}_n^\top \boldsymbol{c}_n)$ are defined by $Y_{ni}(\boldsymbol{z}^\top \mathfrak{a}) = \sum_{j=0}^{\infty} \sum_{\nu=1}^{d_n} z_{n\nu} a_{nj}^{(\nu)} \epsilon_{i-j}$ for $\boldsymbol{z} \in \{\boldsymbol{v}, \boldsymbol{w}\}$ and $\mathfrak{a} \in \{\mathfrak{b}, \mathfrak{c}\}$.

Introduce the filtrations $\mathcal{F}_{nk} = \sigma(\epsilon_{ni} : i \leq k)$, $k \geq 1$, $n \in \mathbb{N}$. In Lemma 2 it is shown that $\boldsymbol{D}_{nk}$ can be approximated by a $\mathcal{F}_{nk}$-martingale array with asymptotic covariance parameter $\beta_n^2(\mathfrak{b}, \mathfrak{c})$ defined in Lemma 1, see (48). Denote by $\alpha_n^2(\mathfrak{a}) = \beta_n^2(\mathfrak{a}, \mathfrak{a})$, for $\mathfrak{a} \in \{\mathfrak{b}, \mathfrak{c}\}$, the associated asymptotic variance parameter.

As a preparation, let $\boldsymbol{B}_n(t) = (\boldsymbol{B}_n^{(1)}(t), \boldsymbol{B}_n^{(2)}(t))^\top$, $t \geq 0$, be a two-dimensional mean zero Brownian motion with variance–covariance matrix

$$\begin{pmatrix} \mathrm{Var}(\boldsymbol{B}_n^{(1)}) & \mathrm{Cov}(\boldsymbol{B}_n^{(1)}, \boldsymbol{B}_n^{(2)}) \\ \mathrm{Cov}(\boldsymbol{B}_n^{(1)}, \boldsymbol{B}_n^{(2)}) & \mathrm{Var}(\boldsymbol{B}_n^{(2)}) \end{pmatrix} = \begin{pmatrix} \alpha_n^2(\mathfrak{b}) & \beta_n^2(\mathfrak{b}, \mathfrak{c}) \\ \beta_n^2(\mathfrak{b}, \mathfrak{c}) & \alpha_n^2(\mathfrak{c}) \end{pmatrix}, \qquad n \geq 1. \tag{18}$$

For $n \geq 1$ define the Gaussian processes

$$G_n(t) = \boldsymbol{B}_n^{(1)}(t)\mathbf{1}(t \leq \tau) + [\boldsymbol{B}_n^{(1)}(\tau) + (\boldsymbol{B}_n^{(2)}(t) - \boldsymbol{B}_n^{(2)}(\tau))]\mathbf{1}(t > \tau), \qquad t \geq 0, \tag{19}$$

$$\overline{G}_n(t) = \frac{1}{\sqrt{n}} G_n(tn), \qquad t \in [0, 1].$$

Before the change, $G_n$ is the Brownian motion $\boldsymbol{B}_n^{(1)}$ with variance $\alpha_n^2(\mathfrak{b})$ and after the change it behaves as the Brownian motion $\boldsymbol{B}_n^{(2)}$ with start in $\boldsymbol{B}_n^{(1)}(\tau)$ and variance $\alpha_n^2(\mathfrak{c})$. Further define

$$G_n^0(k) = G_n(k) - \frac{k}{n} G_n(n), \qquad k \leq n, n \geq 1,$$

$$\overline{G}_n^0(t) = \overline{G}_n(t) - t\overline{G}_n(1), \qquad t \in [0, 1].$$

The covariance functions can be explicitly calculated, for example,

$$\mathrm{Cov}(G_n(s), G_n(t)) = \begin{cases} \min(s, t)\alpha_n^2(\mathfrak{b}), & s, t \leq \tau \text{ or } s \leq \tau < t \text{ or } t \leq \tau < s, \\ \min(s - \tau, t - \tau)\alpha_n^2(\mathfrak{c}), & \tau \leq s, t. \end{cases},$$

but these expressions are not relevant. For further reference, denote

$$\overline{c}_n^0(s, t) = \mathrm{Cov}(\overline{G}_n^0(s), \overline{G}_n^0(t)), \qquad s, t \in [0, 1]. \tag{20}$$

### 3.2. Change-point Gaussian approximations

Closely related to the CUSUM procedures are the following càdlàg processes: Define

$$\mathcal{D}_n(t) = n^{-1/2} \boldsymbol{v}_n^\top (\boldsymbol{S}_{n, \lfloor nt \rfloor} - \lfloor nt \rfloor \mathrm{E}(\boldsymbol{S}_{nn}))\boldsymbol{w}_n, \qquad t \in [0, 1], n \geq 1,$$

and the introduce the associated bridge process

$$\mathcal{D}_n^0(t) = \mathcal{D}_n\left(\frac{\lfloor nt \rfloor}{n}\right) - \frac{\lfloor nt \rfloor}{n} \mathcal{D}_n(1), \qquad t \in [0, 1].$$

Observe that its expectation is $\mathrm{E}(\mathcal{D}_n^0(k/n)) = \frac{1}{\sqrt{n}}\left(\sum_{i=1}^k \sigma_n^2[i] - \frac{k}{n}\sum_{i=1}^n \sigma_n^2[i]\right)$, and vanishes, if $\sigma_n^2[1] = \cdots = \sigma_n^2[n]$. But a non-constant series $\sigma_n^2[i]$, $i \in \{1, \ldots, n\}$, may lead to $\mathrm{E}(\mathcal{D}_n^0(k/n)) \neq 0$. This particularly holds for the change-point model. Our results show that $\mathcal{D}_n(t)$ ($\mathcal{D}_n^0(t)$) can be approximated by a Brownian (bridge) process and lead to a FCLT under weak regularity conditions, and the same holds true for weighted version of these càdlàg processes for nice weighting functions $g$.

Define for $k \geq 1, n \geq 1$,

$$U_{nk} = \boldsymbol{v}_n^\top \boldsymbol{S}_{nk} \boldsymbol{w}_n,$$

$$D_{nk} = U_{nk} - \mathrm{E}(U_{nk}) = \boldsymbol{v}_n^\top (\boldsymbol{S}_{nk} - \mathrm{E}(\boldsymbol{S}_{nk}))\boldsymbol{w}_n,$$

and

$$m_n(k) := \mathrm{E}\left(U_{nk} - \frac{k}{n} U_{nn}\right) = \begin{cases} \frac{k(n-\tau)}{n} \Delta_n, & k \leq \tau, \\ \tau \frac{n-k}{n} \Delta_n, & k > \tau. \end{cases}$$

The following theorem extends the results of [35,36] and justifies the proposed tests (12) and (14) when combined with the results of the next section on consistency of the asymptotic variance parameters. This and all subsequent results consider the basic time series model (1), but all results hold for the multivariate linear processes and VARMA models introduced in Section 2 under the conditions discussed there.

**Theorem 1.** *Suppose that $\{\epsilon_{ni} : i \in \mathbb{Z}, n \geq 1\}$ satisfies Assumption (E). Let $\boldsymbol{v}_n, \boldsymbol{w}_n$ be weighting vectors with $\ell_1$-norms satisfying*

$$\|\boldsymbol{v}_n\|_{\ell_1} \|\boldsymbol{w}_n\|_{\ell_1} = O(n^\eta), \qquad \textit{for } 0 \leq \eta \leq (\theta - \theta')/4 \textit{ for some } 0 < \theta' < \theta, \tag{21}$$

*and let* $\mathfrak{b} = \{b_{nj}^{(\nu)}\}$ *and* $\mathfrak{c} = \{c_{nj}^{(\nu)}\}$ *be coefficients satisfying Assumption (D). If the change-point model (9) holds, then, for each n, one may redefine, on a new probability space, the vector time series together with a two-dimensional mean zero Brownian motion* $\{\boldsymbol{B}_n(t) : t \in [0,1]\}$ *with coordinates* $\boldsymbol{B}_n^{(i)}(t)$, $t \in [0,1]$, $i = 1, 2$, *characterized by the covariance matrix (18) associated to the parameters* $\alpha_n^2(\mathfrak{b})$, $\alpha_n^2(\mathfrak{c})$, *assumed to be bounded away from zero, and* $\beta_n^2(\mathfrak{b}, \mathfrak{c})$, *such that for some constant* $C_n$ *the following assertions hold true almost surely:*

(i) $\|\boldsymbol{D}_{nt} - \boldsymbol{B}_n(t)\|_2 \le C_n t^{1/2-\lambda}$, $t > 0$.

(ii) $\max_{1\le k<n} \|\boldsymbol{D}_{nk} - \frac{k}{n}\boldsymbol{D}_{nn} - [\boldsymbol{B}_{nk} - \frac{k}{n}\boldsymbol{B}_n(n)]\|_2 \le 2C_n n^{1/2-\lambda}$, $n \ge 1$.

(iii) $\max_{1\le k<n} \frac{1}{\sqrt{n}} |D_{nk} - \frac{k}{n} D_{nn} - G_n^0(k)| \le 6\sqrt{2} C_n n^{-\lambda}$, $n \ge 1$.

(iv) $\left|\max_{1\le k<n} \frac{1}{\sqrt{n}} |D_{nk} - \frac{k}{n} D_{nn}| - |\max_{1\le k<n} \frac{1}{\sqrt{n}} |G_n^0(k)|\right| \le 6\sqrt{2} C_n n^{-\lambda}$, $n \ge 1$.

(v) $\max_{1\le k<n} \frac{1}{\sqrt{n}} |U_{nk} - \frac{k}{n} U_{nn} - [m_n(k) + G_n^0(k)]| \le 6\sqrt{2} C_n n^{-\lambda}$, $n \ge 1$.

(vi) $\left|\max_{1\le k<n} \frac{1}{\sqrt{n}} |U_{nk} - \frac{k}{n} U_{nn}| - \max_{1\le k<n} \frac{1}{\sqrt{n}} |m_n(k) + G_n^0(k)|\right| \le 6\sqrt{2} C_n n^{-\lambda}$, $n \ge 1$.

*If* $C_n n^{-\lambda} = o(1)$, *then we also have*

(vii) $\sup_{t\in[0,1]} \left|\mathcal{D}_n(t) - [\mu_n(t) + \overline{G}_n(\lfloor nt \rfloor/n)]\right| = o(1)$, *a.s., as* $n \to \infty$,

(viii) $\sup_{t\in[0,1]} \left|\mathcal{D}_n^0(t) - [\mu_n(t) + \overline{G}_n^0(\lfloor nt \rfloor/n)]\right| = o(1)$, *a.s., as* $n \to \infty$,

*where* $\mu_n(t) = \lfloor nt \rfloor/n(1 - \tau/n)\Delta_n \mathbf{1}(t \le \tau/n) + \tau/n(1 - \lfloor nt \rfloor/n)\Delta_n \mathbf{1}(t > \tau/n)$, $t \in [0,1]$. *Further, provided the weight function g satisfies (13), the corresponding above assertions hold in probability, if* $C_n n^{-\lambda} = o(1)$. *Especially,*

$$\max_{1\le k<n} \frac{1}{\sqrt{n} g(k/n)} \left| U_{nk} - \frac{k}{n} U_{nn} - \left[m_n(k) + G_n^0(k)\right]\right| = o_{\Pr}(1) \tag{22}$$

*and*

$$\left|\max_{1\le k<n} \frac{1}{\sqrt{n} g(k/n)} \left| U_{nk} - \frac{k}{n} U_{nn}\right| - \max_{1\le k<n} \frac{1}{\sqrt{n} g(k/n)} \left|m_n(k) + G_n^0(k)\right|\right| = o_{\Pr}(1). \tag{23}$$

**Remark 1.** Provided the original probability space, $(\Omega, \mathcal{A}, \Pr)$, is rich enough to carry an additional uniform random variable, the strong approximation results of Theorem 1 can be constructed on $(\Omega, \mathcal{A}, \Pr)$.

When there is a change, the drift term $m_n$ yields the consistency of the test.

**Theorem 2.** *Under the assumptions of Theorem 1 and (11),* $\max_{1\le k<n} \frac{1}{\sqrt{n}} \left|U_{nk} - \frac{k}{n} U_{nn}\right| \to \infty$, $n \overset{\Pr}{\to} \infty$.

Note that Theorem 1 holds without the conditions (10) and (11). To discuss conditions of detectability of a change, observe that the drift of the approximating Gaussian process in (22) is given by

$$H_n(k/n) = H_n(k/n; \tau/n, \Delta_n, g) = \sqrt{n}\Delta_n \left[\frac{k(n-\tau)}{n^2 g(k/n)} \mathbf{1}(k \le \tau) + \tau \frac{(n-k)}{n^2 g(k/n)} \mathbf{1}(k > \tau)\right].$$

If this function is asymptotically constant, especially if $\Delta_n \ne 0$ for all $n$ but $\sqrt{n}\Delta_n = o(1)$ (which implies $|\alpha_n^2(\mathfrak{b}) - \alpha_n^2(\mathfrak{c})| = o(1)$ by (8) and Lemma 1) and $\tau/n \to \vartheta \in (0,1)$, then the change is asymptotically not detectable, since the asymptotic law is the same as under the null hypothesis. Now assume $\tau/n \to \vartheta$. A change $\vartheta$ located in a measurable set $A \subset (0,1)$ with positive Lebesgue measure is detectable and changes the asymptotic law, if $H_n \to h^*$, $n \to \infty$, for some function $h^* \ne 0$ on $A$, since then the asymptotic law is given by $\sup_{0<t<1} |[h^*(t) + B^0(t)]/g(t)|$, or if $H_n \overset{\Pr}{\to} \infty$, $n \to \infty$, *on* $[\vartheta, 1)$, cf. Theorem 2. The case $H_n \to h^*$ corresponds to a local alternative such as $\boldsymbol{\Sigma}_{n1} = \boldsymbol{\Sigma}_{n0} + \boldsymbol{\Delta}_n/\sqrt{n}$ for some $d_n \times d_n$ matrix $\boldsymbol{\Delta}_n$ such that $\lim_{n\to\infty} \boldsymbol{v}_n^\top \boldsymbol{\Delta}_n \boldsymbol{w}_n$ exists. For example, if in the spiked covariance model (3) a new local spike term of the form $n^{-1/2}\lambda_{r+1}\boldsymbol{u}_{n,r+1}$ appears after the change-point, then $\boldsymbol{\Delta}_n = \lambda_{r+1}\boldsymbol{u}_{n,r+1}$ and $\Delta_n = \lambda_{r+1}\boldsymbol{v}_n^\top \boldsymbol{u}_{n,r+1}\boldsymbol{w}_n^\top \boldsymbol{u}_{n,r+1}$. Condition (11) is then satisfied, if the weighting vectors are not asymptotically orthogonal to the direction of the new spike.

Observe that $H_n(\cdot, \tau, \Delta_n; g)$ is linear in $\Delta_n = \boldsymbol{v}_n^\top(\boldsymbol{\Sigma}_{n0} - \boldsymbol{\Sigma}_{n1})\boldsymbol{w}_n$. Clearly, $|\Delta_n|$ is maximized if $\boldsymbol{v}_n = \boldsymbol{w}_n$ is a leading eigenvector of $\boldsymbol{\Sigma}_{n0} - \boldsymbol{\Sigma}_{n1}$. This can be seen from the spectral decomposition $\boldsymbol{\Delta}_n = \sum_{i=1}^s \phi_{ni}\boldsymbol{\delta}_{ni}\boldsymbol{\delta}_{ni}^\top$, where $\boldsymbol{\delta}_{ni}$ are the eigenvectors and $\phi_{ni}$ the eigenvalues. When there is no knowledge about the change, e.g., in terms of the $\phi_{ni}$ and/or $\boldsymbol{\delta}_{ni}$ or in terms of the model coefficients $c_{nj}^{(\nu)}$, it makes sense to select $\boldsymbol{v}_n, \boldsymbol{w}_n$ from a known basis or as leading (sparse) eigenvectors of $\boldsymbol{\Sigma}_{n0}$, estimated from a learning sample, in order to obtain a procedure which is capable to react, if the dominant part of the eigenstructure of the covariance matrix changes. Clearly, a change in the orthogonal complement of chosen projection vectors is not detectable. This can be avoided by considering, in addition, random projection(s).

For the CUSUM statistics based on the centered time series we have the following approximation result.

**Theorem 3.** *Let the original probability space be rich enough to carry an additional uniform random variable. Assume the conditions of Theorem 1 and the strengthened decay condition* $\sup_{n\ge1} \max_{1\le\nu\le d_n} |c_{nj}^{(\nu)}| \ll (j \vee 1)^{-1-\theta}$ *for some* $\theta > 0$ *hold.*

*Suppose that the vector time series is centered at the sample averages $\widehat{\mu}_\nu = \frac{1}{n}\sum_{i=1}^n Y_{ni}^{(\nu)}$, before applying the CUSUM procedures, leading to the statistics $\widetilde{C}_n(g)$ and $\widetilde{T}_n(g)$. Then assertions (i) and (ii) of Theorem 1 hold true with an additional error term $o_{\Pr}(n^{1/2})$ and (iii)-(vi) with an additional $o_{\Pr}(1)$ term. Finally, (vii) and (viii) hold in probability, if $C_n n^{-\lambda} = o(1)$.*

The above theorems assume that the projection vectors $\boldsymbol{v}_n$ and $\boldsymbol{w}_n$ have uniformly bounded $\ell_1$-norm. When standardizing by a homogeneous estimator $\widehat{\alpha}_n = \widehat{\alpha}_n(\boldsymbol{v}_n, \boldsymbol{w}_n)$, i.e. satisfying

$$\widehat{\alpha}_n(x\boldsymbol{v}_n, y\boldsymbol{w}_n) = xy\widehat{\alpha}_n(\boldsymbol{v}_n, \boldsymbol{w}_n) \tag{24}$$

for all $x, y > 0$, one can relax the conditions on the projections $\boldsymbol{v}_n, \boldsymbol{w}_n$.

**Theorem 4.** *Suppose that $\{\epsilon_{ni} : i \in \mathbb{Z}, n \geq 1\}$ satisfies Assumption (E). Assume that*

$$\sup_{n\geq 1} d_n^{-1/2}\|\boldsymbol{v}_n\|_{\ell_2}, \sup_{n\geq 1} d_n^{-1/2}\|\boldsymbol{w}_n\|_{\ell_2} < \infty \tag{25}$$

*or there are non-decreasing sequences $\{a_n\}, \{b_n\} \subset (0, \infty)$ with*

$$\sup_{n\geq 1} a_n^{-1}\|\boldsymbol{v}_n\|_{\ell_1}, \sup_{n\geq 1} b_n^{-1}\|\boldsymbol{w}_n\|_{\ell_1} < \infty. \tag{26}$$

*Suppose that the estimator $\widehat{\alpha}_n = \widehat{\alpha}_n(\boldsymbol{v}_n, \boldsymbol{w}_n)$ used by $T_n(g; \boldsymbol{v}_n, \boldsymbol{w}_n)$ is ratio consistent and homogeneous. Further, let $\mathfrak{b} = \{b_{nj}^{(\nu)}\}$ and $\mathfrak{c} = \{c_{nj}^{(\nu)}\}$ be coefficients satisfying Assumption (D). If the change-point model (9) holds, then, under the construction of Theorem 1 with $C_n n^{-\lambda} = o(1)$, (vi) holds and we have for any weight function $g$ satisfying (13)*

$$\left| T_n(g) - \max_{1\leq k<n} \frac{1}{g(k/n)} \left| \frac{m_n(k)}{\sqrt{n}} + \overline{B}_n^0(k/n) \right| \right| = o_{\Pr}(1), \tag{27}$$

*where $\overline{B}_n^0(t) = \alpha_n^{-1}(\mathfrak{b})\overline{G}_n^0(t)$, $t \in [0, 1]$.*

By Theorem 1, statistical properties of the CUSUM statistic $C_n(g)$ can be approximated by those of $\max_{1\leq k<n} \frac{1}{\sqrt{n}g(k/n)} \left|m_n(k) + G_n^0(k)\right|$. In view of Theorem 4, for the standardized CUSUM statistic $T_n(g)$ one replaces $G_n^0$ by a process which is a Brownian bridge with covariance function $\min(s, t) - st$ up to $\tau$ and $(\min(s, t) - st)\alpha_n^2(\mathfrak{c})/\alpha_n^2(\mathfrak{b})$ after the change. Especially, under the null hypothesis $H_0$ of no change, we have $m_n(k) = 0$, for all $k$ and $n$, and $|\alpha_n^2(\mathfrak{b}) - \alpha^2(\mathfrak{c})| = o(1)$ by (8) and Lemma 1. Then the asymptotics of the change-point procedures is governed by a standard Brownian bridge. Theorems 1, 3 and 4 (under the strengthened decay condition) imply FCLTs.

**Theorem 5** (*FCLT*)**.** *If $\beta_n^2(\mathfrak{b}, \mathfrak{c}) \to \beta^2(\mathfrak{b}, \mathfrak{c})$, $\alpha_n^2(\mathfrak{a}) \to \alpha^2(\mathfrak{a}) > 0$ for $\mathfrak{a} \in \{\mathfrak{b}, \mathfrak{c}\}$, $\Delta_n \to \Delta > 0$ and $\tau/n \to \vartheta \in (0, 1)$, as $n \to \infty$, then under the conditions of Theorem 1 (viii) or Theorem 4 it holds*

$$\mathcal{D}_n^0 \Rightarrow \mu + \overline{G}^0, \qquad n \to \infty,$$

*with $\mu(t) = t(1-\vartheta)\Delta\mathbf{1}(t \leq \vartheta) + \vartheta(1-t)\Delta\mathbf{1}(t > \vartheta)$, $t \in [0, 1]$, in the Skorohod space $D[0, 1]$, for some Gaussian bridge-type process $\overline{G}^0$ defined on $[0, 1]$, which is a Brownian bridge with variance parameter $\alpha^2(\mathfrak{b})$ under $H_0$. Further, if $\boldsymbol{v}_n, \boldsymbol{w}_n$ are weighting vectors satisfying (21), (25) or (26) and if the constructions of Theorems 1 and 4, respectively, hold with $C_n n^{-\lambda} = o(1)$, then for any Lipschitz weight function $g$ which satisfies (13) we have*

$$T_n(g), \widetilde{T}_n(g) \Rightarrow \sup_{0<t<1} \frac{|\mu(t) + \overline{B}^0(t)|}{g(t)}, \qquad n \to \infty, \quad \text{in } D[0, 1],$$

*where $\overline{B}^0 = \overline{G}^0/\alpha(\mathfrak{b})$. Especially, under the null hypothesis the limiting law is given by $\sup_{0<t<1}|B^0(t)|/g(t)$ for some standard Brownian bridge $B^0$.*

### 3.3. Multivariate CUSUM approximation

Let us now consider $L = L_n \in \mathbb{N}$ CUSUM statistics $\boldsymbol{C}_n(g) = (C_{n1}(g), \ldots, C_{nL_n}(g))^\top$ where

$$C_{nj}(g) = C_n(\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj}; g) = \max_{1\leq k<n} \frac{1}{\sqrt{n}g(k/n)} \left| \boldsymbol{v}_{nj}^\top \left( \boldsymbol{S}_{nk} - \frac{k}{n}\boldsymbol{S}_{nn} \right) \boldsymbol{w}_{nj} \right|,$$

$j \in \{1, \ldots, L_n\}$, defined for $L_n$ pairs $(\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj})$, $j \in \{1, \ldots, L_n\}$, of projection vectors. When using no weights, i.e., $g(x) = 1$, $x \in [0, 1]$, the corresponding quantities are denoted $\boldsymbol{C}_n = (C_{n1}, \ldots, C_{nL_n})^\top$.

Let $\boldsymbol{B}_n(t) = (\boldsymbol{B}_n^{(1)}(t), \ldots, \boldsymbol{B}_n^{(2L_n)})^\top$, $t \geq 0$, be a $2L_n$-dimensional mean zero Brownian motion with covariance matrix

$$\boldsymbol{\Sigma}_n^{\boldsymbol{B}} = \left(\boldsymbol{\Sigma}_{nij}^{\boldsymbol{B}}\right)_{\substack{1\leq i\leq L_n \\ 1\leq j\leq L_n}} \tag{28}$$

with blocks

$$\boldsymbol{\Sigma}^{\boldsymbol{B}}_{nij} = \begin{pmatrix} \beta_n^2(\mathfrak{b}, i, \mathfrak{b}, j) & \beta_n^2(\mathfrak{b}, i, \mathfrak{c}, j) \\ \beta_n^2(\mathfrak{c}, i, \mathfrak{b}, j) & \beta_n^2(\mathfrak{c}, i, \mathfrak{c}, j) \end{pmatrix}, \qquad 1 \le i, j \le L_n,$$

where, for brevity, $\beta_n^2(\mathfrak{b}, i, \mathfrak{c}, j) = L_n^{-\iota}\beta_n^2(\mathfrak{b}, \boldsymbol{v}_{ni}, \boldsymbol{w}_{ni}, \mathfrak{c}, \boldsymbol{v}_{nj}, \boldsymbol{w}_{nj})$ with $\iota = \mathbf{1}(L_n \to \infty)$, $i, j \in \{1, \dots, L\}$. Also put $\alpha_n^2(\mathfrak{a}, i) = L_n^{-\iota}\beta_n^2(\mathfrak{a}, \boldsymbol{v}_{ni}, \boldsymbol{w}_{ni}, \mathfrak{a}, \boldsymbol{v}_{ni}, \boldsymbol{w}_{ni})$, $\mathfrak{a} \in \{\mathfrak{b}, \mathfrak{c}\}$, $i \in \{1, \dots, L\}$, see (48). Define the processes

$$\boldsymbol{G}_n(t) = \boldsymbol{B}_n^{(1)}\mathbf{1}(t \le \tau) + [\boldsymbol{B}_n^{(1)}(\tau) + (\boldsymbol{B}_n^{(2)}(t) - \boldsymbol{B}_n^{(2)}(\tau))]\mathbf{1}(t > \tau), \qquad t \ge 0,$$

$$\boldsymbol{G}_n^0(k) = \boldsymbol{G}_n(k) - \frac{k}{n}\boldsymbol{G}_n(n), \qquad k \le n, n \ge 1,$$

where $\boldsymbol{B}_n^{(1)}(t) = \big(\boldsymbol{B}_{n,2j-1}(t)\big)_{j=1}^{L_n}$, $\boldsymbol{B}_n^{(2)}(t) = \big(\boldsymbol{B}_{n,2j}(t)\big)_{j=1}^{L_n}$ and $\boldsymbol{G}_n^0(k) = \big(\boldsymbol{G}_{nj}^0(k)\big)_{j=1}^{2L_n}$.

**Theorem 6.** *Suppose that $\{\epsilon_{ni} : i \in \mathbb{Z}, n \ge 1\}$ satisfies Assumption (E). Let $\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj}$, $j \in \{1, \dots, L_n\}$, be weighting vectors satisfying (21) uniformly in j, and let $\mathfrak{b} = \{b_{nj}^{(\nu)}\}$ and $\mathfrak{c} = \{c_{nj}^{(\nu)}\}$ be coefficients satisfying Assumption (D). Then, under the change-point model (9), one can redefine, for each n, on a new probability space, the vector time series together with a $2L_n$-dimensional mean zero Brownian motion $\boldsymbol{B}_n = (\boldsymbol{B}_{nj})_{j=1}^{2L_n}$ with covariance function given by (28), such that*

$$\left\| L_n^{-\iota/2}\boldsymbol{C}_n - \left( \max_{1 \le k < n} \frac{1}{L_n^{\iota/2}\sqrt{n}} \left| m_{nj}(k) + \boldsymbol{G}_{nj}^0(k) \right|_2 \right)_{j=1}^{L_n} \right\|_\infty \le 6\sqrt{2}C_n.n^{-\lambda}, \tag{29}$$

*and for a weight function g satisfying (13), for any $\delta > 0$*

$$\max_{j \le L_n} \Pr\left( \left| L_n^{-1/2} C_n(\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj}; g) - \max_{1 \le k < n} \frac{1}{\sqrt{n g(k/n)}} |m_{nj}(k) - \boldsymbol{G}_{nj}^0(k)| \right| > \delta \right) = o(1), \tag{30}$$

*where $m_{nj}(k) = \frac{k(n-\tau)}{n}\Delta_n(j)\mathbf{1}(k \le \tau) + \tau\frac{n-k}{n}\Delta_n(j)\mathbf{1}(\tau < k \le n)$ with $\Delta_n(j) = (\boldsymbol{v}_{nj}^\top \boldsymbol{\Sigma}_{n0}\boldsymbol{w}_{nj} - \boldsymbol{v}_{nj}^\top \boldsymbol{\Sigma}_{n1}\boldsymbol{w}_{nj})$, $j \in \{1, \dots, L_n\}$.*

Observe that under $H_0$ the asymptotic covariance matrix of the approximating process and hence of $\boldsymbol{C}_n$ is given by $\boldsymbol{\Sigma}^{\boldsymbol{B}}_{n,H_0} = \mathrm{Var}(\boldsymbol{B}_n^{(1)}) = \boldsymbol{\Sigma}^{\boldsymbol{B}}_n(\mathfrak{b})$, whose diagonal is given by $\alpha_n^2(\mathfrak{b}, 1), \dots, \alpha_n^2(\mathfrak{b}, L_n)$ and off-diagonal elements by $\sigma_n^2(\mathfrak{b}, i, \mathfrak{b}, j)$, $1 \le i \ne j \le L_n$. For fixed $L$ the results of the next section show that $\boldsymbol{\Sigma}^{\boldsymbol{B}}_n(\mathfrak{b})$ can be estimated consistently, providing a heuristic justification for the test (15) when $\boldsymbol{\Sigma}^{\boldsymbol{B}}_n(\mathfrak{b})$ is regular.

### *3.4. Full-sample and stopped-sample estimation of $\alpha_n^2(\mathfrak{b})$ and $\beta_n^2(\mathfrak{b}, \mathfrak{c})$*

Let us now discuss how to estimate the parameter $\alpha_n^2(\mathfrak{b})$ for one pair $(\boldsymbol{v}_n, \boldsymbol{w}_n)$ of projection vectors, which is used in the change-point test statistic for standardization, and the asymptotic covariance parameters $\beta_n^2(j, k) = \beta_n^2(\mathfrak{b}, \boldsymbol{v}_{nj}, \boldsymbol{w}_{nj}, \mathfrak{b}, \boldsymbol{v}_{nk}, \boldsymbol{w}_{nk})$ for two pairs $(\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj})$ and $(\boldsymbol{v}_{nk}, \boldsymbol{w}_{nk})$, which arise in the multivariate test for a set of projections. If there is no change, one may use the proposal of [35]. But under a change these estimators are inconsistent. The common approach is therefore to use a learning sample for estimation. Alternatively, one may estimate the change-point and use the data before the change. The consistency of that approach follows quite easily when establishing a uniform weak of large numbers of the sequential (process) version of the estimators which uses the first $k$ observations, $\boldsymbol{Y}_{n1}, \dots, \boldsymbol{Y}_{nk}$, where $k$ is a fraction of the sample size $n$ so that $k = \lfloor nu \rfloor$ for $u \in (0, 1]$:

Fix $0 < \varepsilon < 1$ and define for $u \in [\varepsilon, 1]$

$$\widehat{\alpha}_n^2(u) = \widehat{\Gamma}_n(u; 0) + 2\sum_{h=1}^m w_{mh}\widehat{\Gamma}_n(u; h),$$

where

$$\widehat{\Gamma}_n(u; h) = \frac{1}{\lfloor nu \rfloor} \sum_{i=1}^{\lfloor nu \rfloor - h} [\boldsymbol{v}_n^\top \boldsymbol{Y}_{ni}\boldsymbol{w}_n^\top \boldsymbol{Y}_{ni} - \widehat{c}_{\lfloor nu \rfloor}][\boldsymbol{v}_n^\top \boldsymbol{Y}_{n,i+|h|}\boldsymbol{w}_n^\top \boldsymbol{Y}_{n,i+|h|} - \widehat{c}_{\lfloor nu \rfloor}],$$

for $|h| \le m$, with $\widehat{c}_{\lfloor nu \rfloor} = \lfloor nu \rfloor^{-1}\sum_{j=1}^{\lfloor nu \rfloor} \boldsymbol{v}_n^\top \boldsymbol{Y}_{nj}\boldsymbol{w}_n^\top \boldsymbol{Y}_{nj}$. The estimators $\widehat{\beta}_n^2(j, k)$ and $\widehat{\Gamma}_n(u; h, j, k)$, $1 \le j, k \le K$, corresponding to two pairs of projection vectors, are defined analogously, i.e.,

$$\widehat{\beta}_n^2(j, k) = \widehat{\Gamma}_n(u; 0, j, k) + 2\sum_{h=1}^m w_{mh}\widehat{\Gamma}_n(u; h, j, k) \tag{31}$$

with

$$\widehat{\Gamma}_n(u; h, j, k) = \frac{1}{\lfloor nu \rfloor} \sum_{i=1}^{\lfloor nu \rfloor - h} [\boldsymbol{v}_{nj}^\top \boldsymbol{Y}_{ni}\boldsymbol{w}_{nj}^\top \boldsymbol{Y}_{ni} - \widehat{c}_{\lfloor nu \rfloor, j}][\boldsymbol{v}_{nk}^\top \boldsymbol{Y}_{n,i+|h|}\boldsymbol{w}_{nk}^\top \boldsymbol{Y}_{n,i+|h|} - \widehat{c}_{\lfloor nu \rfloor, k}] \tag{32}$$

for $1 \le i, j \le L$ with $\widehat{c}_{\lfloor nu \rfloor, j} = \lfloor nu \rfloor^{-1}\sum_{i=1}^{\lfloor nu \rfloor} \boldsymbol{v}_{nj}^\top \boldsymbol{Y}_{ni}\boldsymbol{w}_{nj}^\top \boldsymbol{Y}_{ni}$, $j \in \{1, \dots, L\}$.

The weights are often defined through a kernel function $w(x)$ via $w_{mh} = w(h/b_m)$ for some bandwidth parameter $b_m$. For a brief discussion of common choices see [35].

The following theorem establishes the uniform law of large numbers. Especially, it shows that $\widehat{\alpha}_n^2(\ell/n)$ is consistent for $\alpha^2(\mathfrak{b})$ if $\ell \le \tau$, whereas for $\ell > \tau$ a convex combination of $\alpha^2(\mathfrak{b})$ and $\alpha^2(\mathfrak{c})$ is estimated. A similar result applies to the estimator of the asymptotic covariance parameter.

**Theorem 7.** *Assume that $m \to \infty$ with $m^2/n = o(1)$, as $n \to \infty$, and the weights $\{w_{mh}\}$ satisfy*

(i) $w_{mh} \to 1$, *as* $m \to \infty$, *for all* $h \in \mathbb{Z}$, *and*
(ii) $0 \le w_{mh} \le W < \infty$, *for some constant* $W$, *for all* $m \ge 1$, $h \in \mathbb{Z}$.

*If the innovations $\epsilon_{ni} = \epsilon_i$ are i.i.d. with $\mathrm{E}(\epsilon_1^8) < \infty$, $c_{nj}^{(\nu)} = c_j^{(\nu)}$, for all $j$ and $n \ge 1$, satisfy the decay condition*

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

*for some $\delta > 0$, and $\boldsymbol{v}, \boldsymbol{w} \in \ell_1$, then under the change-in-coefficients model (9) with $\tau = \lfloor n\vartheta \rfloor$, $\vartheta \in (0,1)$, it holds for any $0 < \varepsilon < \vartheta$*

$$\sup_{u \in [\varepsilon, 1]} |\widehat{\alpha}_n^2(u) - \alpha^2(u)| \stackrel{\Pr}{\to} 0,$$

*as $n \to \infty$, where $\alpha^2(u) = \alpha^2(u; \mathfrak{b}, \mathfrak{c}) = \mathbf{1}(u \le \vartheta)\alpha^2(\mathfrak{b}) + \mathbf{1}(u > \vartheta)\big((\vartheta/u)\alpha^2(\mathfrak{b}) + (1-\vartheta/u)\alpha^2(\mathfrak{c})\big)$, for $u \in [\varepsilon, 1]$. Further,*

$$\sup_{u \in [\varepsilon, 1]} |\widehat{\beta}_n^2(u, j, k) - \beta^2(u, j, k)| \stackrel{\Pr}{\to} 0,$$

*where for $u \in [\varepsilon, 1]$ $\beta^2(u, j, k) = \beta^2(u; j, k, \mathfrak{b}, \mathfrak{c}) = \mathbf{1}(u \le \vartheta)\beta^2(\mathfrak{b}, j, k) + \mathbf{1}(u > \vartheta)\big((\vartheta/u)\beta^2(\mathfrak{b}, j, k) + (1-\vartheta/u)\beta^2(\mathfrak{c}, j, k)\big)$, for $1 \le j, k \le L$, as defined in Lemma 1.*

Let us now suppose we are given a consistent estimator $\widehat{\tau}_n$ of the unknown change-point; in the next section we make a concrete proposal. In order to estimate the parameter $\alpha^2(\mathfrak{b})$ it is natural to use the above estimator using all observations classified by the estimator as belonging to the pre-change period. This means, we estimate $\alpha^2(\mathfrak{b})$ by $\widehat{\alpha}^2_{\widehat{\tau}_n}$. The following result shows that this estimator is consistent under weak conditions.

**Theorem 8.** *Suppose that $\widehat{\tau}_n$ is an estimator of $\tau$ satisfying $\widehat{\tau}_n/n \in [\varepsilon, 1]$ a.s. and $\left|\frac{\widehat{\tau}_n}{n} - \vartheta\right| \stackrel{\Pr}{\to} 0$, as $n \to \infty$. Then*

$$|\widehat{\alpha}^2_{\widehat{\tau}_n} - \alpha^2(\mathfrak{b})| \stackrel{\Pr}{\to} 0, \qquad n \to \infty.$$

## 4. Change-point estimation

In view of the change-point test statistic studied in the previous section, it is natural to estimate the change-point $\widehat{\tau}_n$ by

$$\widehat{\tau}_n = \underset{1 \le k < n}{\operatorname{argmax}} |\widehat{\mathcal{U}}_n(k)|, \qquad \widehat{\mathcal{U}}_n(k) = \frac{1}{g(k/n)n}\left(U_{nk} - \frac{k}{n}U_{nn}\right), \qquad 1 \le k \le n, n \ge 1.$$

(By convention, $\operatorname{argmax}_{x \in \mathcal{D}} f(x)$ denotes the smallest maximizer of some function $f : \mathcal{D} \to \mathbb{R}$.)

The expectation $m_n(k)$ of $U_{nk} - \frac{k}{n}U_{nn}$ is a function of $\Delta_n$, and we assume that the limit

$$\Delta = \lim_{n \to \infty} \Delta_n, \tag{33}$$

exists. To proceed, we need further notation. Put

$$\mathcal{U}_n(k) = \mathrm{E}(\widehat{\mathcal{U}}_n(k)) = \begin{cases} \frac{k(n-\tau)}{g(k/n)n^2}\Delta_n, & k \le \tau, \\ \tau\frac{n-k}{g(k/n)n^2}\Delta_n, & k > \tau, \end{cases} \tag{34}$$

and introduce the associated rescaled functions

$$\widehat{u}_n(t) = \widehat{\mathcal{U}}_n(\lfloor nt \rfloor), \qquad t \in [0,1], \tag{35}$$

$$u_n(t) = \mathcal{U}_n(\lfloor nt \rfloor), \qquad t \in [0,1], \tag{36}$$

and

$$u(t) = \frac{t}{g(t)}(1-\vartheta)\Delta\mathbf{1}(t \le \vartheta) + \vartheta\frac{1-t}{g(t)}\Delta\mathbf{1}(t > \vartheta), \qquad t \in [0,1]. \tag{37}$$

If $g = 1$, then for $\Delta > 0$ the function $u(t)$ is strictly increasing on $[0, \vartheta]$ and strictly decreasing on $[\vartheta, 1]$, and for $\Delta < 0$ the same holds for $|u(t)|$. The same applies for any weight function $g$ such that

$$g \text{ is continuous, } t/g(t) \text{ increasing on } [0, \vartheta] \text{ and } (1-t)/g(t) \text{ decreasing on } [\vartheta, 1]. \tag{38}$$

Obviously, this holds for a large class of functions $g$ whatever the value of the true change-point. Hence, we expect that the maximizers of $|u_n(t)|$, $t \in [0, 1]$, and its estimator $|\widehat{u}_n(t)|$, $t \in [0, 1]$, converge to the true change-point $\vartheta$. But the maximizers $\widehat{\tau}_n$ of $|\widehat{\mathcal{U}}_n|$ and $\widehat{t}_n$ of $|\widehat{u}_n|$ are related by

$$\widehat{\tau}_n = \underset{1 \le k \le n}{\operatorname{argmax}} \widehat{\mathcal{U}}_n(k) = \underset{1 \le k \le n}{\operatorname{argmax}} \widehat{u}_n(k/n) = n \underset{t \in \{1/n, \dots, 1\}}{\operatorname{argmax}} \widehat{u}_n(t) = n\widehat{t}_n \tag{39}$$

Therefore, since $\widehat{u}_n$ is constant on $[k/n, (k+1)/n)$, $k \in \{1, \dots, n-1\}$ and vanishes on $[0, 1/n)$, $\widehat{t}_n \xrightarrow{\Pr} \vartheta$, as $n \to \infty$ implies $\frac{\lfloor \widehat{\tau}_n \rfloor}{n} \xrightarrow{\Pr} \vartheta$, as $n \to \infty$.

A martingale approximation and Doob's inequality provide the following uniform convergence.

**Theorem 9.** *Let $g$ be a weight function satisfying (13) and (38). If (33) holds, then*

$$\max_{1 \le k < n} |\widehat{\mathcal{U}}_n(k) - \mathcal{U}_n(k)| \xrightarrow{\Pr} 0, \qquad n \to \infty, \tag{40}$$

$$\sup_{t \in [0,1]} |\widehat{u}_n(t) - u(t)| \xrightarrow{\Pr} 0, \qquad n \to \infty. \tag{41}$$

The consistency of the change-point estimator $\widehat{\tau}_n$ follows now easily from the above results.

**Theorem 10.** *Under the assumptions of Theorem 9 and the change-point alternative model (9) with $\tau = \lfloor n\vartheta \rfloor$, $\vartheta \in (\varepsilon, 1)$ for some $\varepsilon \in (0, 1)$, we have*

$$\frac{\widehat{\tau}_n}{n} \xrightarrow{\Pr} \vartheta, \qquad n \to \infty.$$

## 5. Simulations

To investigate the statistical performance of the change-point tests a change from a family of AR($\rho_\nu$) series to a family of (shifted) MA($r$) series, which are, at lag 0, independent, was examined: We assume that these series, $Y_{ni}^{(\nu)}$, are defined as follows. Fix $r \in \mathbb{N}$ and let

$$\text{pre-change } (i \le \tau)\text{: } Y_{ni}^{(\nu)} = \rho_\nu Y_{n,i-1}^{(\nu)} + \epsilon_{i-1}, \quad \text{after-change } (i > \tau)\text{: } Y_{ni}^{(\nu)} = \sum_{j=0}^{r} \theta_r^{(\nu)} \epsilon_{i-j-(\nu-1)r},$$

with $\rho_\nu = 0.5\nu/d$, for $\nu \in \{1, \dots, d\}$ and $n \ge 1$, i.i.d. standard normal $\epsilon_t$ and $\theta_j^{(\nu)} = (1 - 0.1 \cdot j)\sqrt{(1-\rho_\nu^2)^{-1}/s_\theta^2}$, $s_\theta^2 = \sum_{k=0}^{4}(1 - 0.1 \cdot k)^2$, $j \in \{0, \dots, r = 4\}$, so that the marginal variances of the $d$ time series do not change. The asymptotic variance parameter, $\alpha_n$, was estimated with lag truncation $m = \lceil n^{1/3} \rceil$ justified by simulations not reported here, using three sampling approaches: (i) Learning sample of size $L = 500$, (ii) full in-sample estimation and (iii) stopped in-sample estimation using the modified rule $\widetilde{\tau}_n = \max(\lfloor n/4 \rfloor, \min(1.15 \cdot \widehat{\tau}_n, n))$. Although this modification may lead to some bias, the actual number of observations was increased, since otherwise the sample size for estimation may be too small.

Both a fixed and a random projection were examined. The case of a fixed projection vector was studied by using $\boldsymbol{v}_n = \boldsymbol{w}_n = (1/d, \dots, 1/d)^\top$. Random projections were generated by drawing from a Dirichlet distribution, such that the projections have unit $\ell_1$ norm and expectation $d^{-1}\mathbf{1}$, in order to study the effect of random perturbations around the fixed projections.

Table 1 provides the rejection rates for $n = 100$ and dimensions $d \in \{10, 100, 200\}$ when the change-point is given by $\tau = \lfloor n\vartheta \rfloor$ with $\vartheta \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$, to study changes within the central 50% of the data as well as early and late changes. First, one can notice that the power is somewhat increasing in the dimension but quickly saturates. The results for stopped-sample and in-sample estimation are quite similar. The unweighted CUSUM procedure has very accurate type I error rate if a learning sample is present, whereas the weighted CUSUM overreacts somewhat under the null hypothesis. For stopped-sample and in-sample estimation the unweighted procedure is conservative, whereas the weighted CUSUM keeps the level quite well with only little overreaction. Although the unweighted CUSUM operates at a smaller significance level, it is more powerful than the weighted procedure when the change occurs in the middle of the sample, but the weighted CUSUM performs better for early changes. The results for a random projection are very similar.

**Table 1**
Simulated power for the sample size $n = 100$ for fixed projection and a random projection for dimension $d = 10, 100, 200$ and different change-point locations. The entries for $\vartheta = 1$ provide simulated type I error rates.

| Fixed projection | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\vartheta$ | 10 | 100 | 200 | $\vartheta$ | 10 | 100 | 200 |
| Method | Unweighted CUSUM | | | | Weighted CUSUM | | | |
| | $\vartheta$ | 10 | 100 | 200 | $\vartheta$ | 10 | 100 | 200 |
| | 0.10 | 0.03 | 0.02 | 0.02 | 0.10 | 0.15 | 0.14 | 0.14 |
| | 0.25 | 0.31 | 0.34 | 0.34 | 0.25 | 0.35 | 0.37 | 0.38 |
| $L = 500$ | 0.50 | 0.70 | 0.75 | 0.77 | 0.50 | 0.55 | 0.59 | 0.61 |
| | 0.75 | 0.39 | 0.46 | 0.46 | 0.75 | 0.33 | 0.40 | 0.38 |
| | 0.90 | 0.07 | 0.08 | 0.08 | 0.90 | 0.09 | 0.08 | 0.09 |
| | 1.00 | 0.05 | 0.06 | 0.05 | 1.00 | 0.09 | 0.11 | 0.10 |
| | 0.10 | 0.14 | 0.14 | 0.09 | 0.10 | 0.79 | 0.98 | 0.99 |
| | 0.25 | 0.79 | 0.88 | 0.88 | 0.25 | 0.86 | 0.93 | 0.92 |
| stopped-sample | 0.50 | 0.90 | 0.93 | 0.93 | 0.50 | 0.70 | 0.72 | 0.71 |
| | 0.75 | 0.29 | 0.30 | 0.29 | 0.75 | 0.17 | 0.17 | 0.18 |
| | 0.90 | 0.03 | 0.02 | 0.02 | 0.90 | 0.04 | 0.03 | 0.04 |
| | 1.00 | 0.02 | 0.01 | 0.01 | 1.00 | 0.07 | 0.07 | 0.08 |
| | 0.10 | 0.14 | 0.13 | 0.11 | 0.10 | 0.79 | 0.98 | 0.99 |
| | 0.25 | 0.79 | 0.87 | 0.86 | 0.25 | 0.86 | 0.93 | 0.92 |
| in-sample | 0.50 | 0.90 | 0.92 | 0.93 | 0.50 | 0.69 | 0.73 | 0.72 |
| | 0.75 | 0.29 | 0.30 | 0.29 | 0.75 | 0.17 | 0.17 | 0.18 |
| | 0.90 | 0.03 | 0.02 | 0.02 | 0.90 | 0.04 | 0.04 | 0.04 |
| | 1.00 | 0.02 | 0.02 | 0.02 | 1.00 | 0.08 | 0.08 | 0.07 |
| Random projection | | | | | | | | |
| Method | Unweighted CUSUM | | | | Weighted CUSUM | | | |
| | $\vartheta$ | 10 | 100 | 200 | $\vartheta$ | 10 | 100 | 200 |
| | 0.10 | 0.02 | 0.02 | 0.02 | 0.10 | 0.14 | 0.14 | 0.14 |
| | 0.25 | 0.32 | 0.36 | 0.36 | 0.25 | 0.36 | 0.39 | 0.39 |
| $L = 500$ | 0.50 | 0.71 | 0.77 | 0.76 | 0.50 | 0.54 | 0.60 | 0.61 |
| | 0.75 | 0.39 | 0.46 | 0.46 | 0.75 | 0.33 | 0.39 | 0.39 |
| | 0.90 | 0.08 | 0.08 | 0.08 | 0.90 | 0.09 | 0.09 | 0.10 |
| | 1.00 | 0.05 | 0.05 | 0.05 | 1.00 | 0.10 | 0.10 | 0.09 |
| | 0.10 | 0.13 | 0.15 | 0.11 | 0.10 | 0.80 | 0.98 | 0.99 |
| | 0.25 | 0.80 | 0.86 | 0.87 | 0.25 | 0.85 | 0.92 | 0.92 |
| stopped-sample | 0.50 | 0.90 | 0.93 | 0.93 | 0.50 | 0.70 | 0.72 | 0.72 |
| | 0.75 | 0.28 | 0.29 | 0.30 | 0.75 | 0.18 | 0.17 | 0.17 |
| | 0.90 | 0.03 | 0.03 | 0.03 | 0.90 | 0.04 | 0.04 | 0.04 |
| | 1.00 | 0.02 | 0.02 | 0.02 | 1.00 | 0.08 | 0.07 | 0.07 |
| | 0.10 | 0.14 | 0.14 | 0.11 | 0.10 | 0.80 | 0.98 | 0.98 |
| | 0.25 | 0.79 | 0.87 | 0.87 | 0.25 | 0.86 | 0.92 | 0.92 |
| in-sample | 0.50 | 0.91 | 0.93 | 0.93 | 0.50 | 0.71 | 0.72 | 0.73 |
| | 0.75 | 0.28 | 0.30 | 0.30 | 0.75 | 0.17 | 0.18 | 0.16 |
| | 0.90 | 0.03 | 0.03 | 0.02 | 0.90 | 0.04 | 0.04 | 0.04 |
| | 1.00 | 0.02 | 0.02 | 0.02 | 1.00 | 0.08 | 0.07 | 0.08 |

## 6. Data example

To illustrate the proposed methods, we analyze $n = 1826$ daily observations of 8 h maxima of ozone concentration collected at $d = 444$ monitors in the U.S.. The data corresponds to the 5-year-period from January 2010 to December 2014. We analyze mean corrected data, see [31], namely residuals obtained after fitting cubic splines to the log-transformed data, in order to correct level and seasonal ups and downs. The aim of the analysis is to check whether there is evidence for a change in the dependence structure over time.

The data of the first year was used to calculate a sparse PCA. We use the method of [11] to get sparse directions $\boldsymbol{v}_i$ instead of [6], since, according to the latter authors, their estimators leading to minimax rates are computationally infeasible, for details see the supplement. This analysis shows that the supports $\mathcal{S}_i = \{j : \boldsymbol{v}_{ij} \neq 0\}$ of the leading six projections, where $\boldsymbol{v}_i = (v_{i1}, \dots, v_{id})^\top$ for $i \in \{1, \dots, 6\}$, correspond to a spatial segmentation which eases interpretation. The data of the years 2011 to 2014, providing the test sample $\boldsymbol{Y}_{n1}, \dots, \boldsymbol{Y}_{nn}$ with $n = 1462$, was now analyzed using the leading directions as projection vectors.

The proposed change-point tests were applied to test for the presence of changes in the (co-) variances $\mathrm{Cov}(\boldsymbol{v}_k^\top \boldsymbol{Y}_{ni}, \boldsymbol{v}_\ell^\top \boldsymbol{Y}_{ni})$, $i \in \{1, \dots, n\}$, for $1 \leq k \leq \ell \leq 7$. The asymptotic variance parameter was estimated using both the full in-sample and stopped-sample approach. Using the algorithm of [25] to compute the distribution function and the associated quantile function of the Kolmogorov distribution there is no evidence for a change.

## 7. Proofs

The proofs are based on martingale approximations, which require several additional results and technical preparations. These results extend and complement the results obtained in [35].

### 7.1. Preliminaries

For an arbitrary array of coefficients $\mathfrak{a} = \{a_{nj}^{(\nu)} : j \geq 0, 1 \leq \nu \leq d_n, n \geq 1\}$ and vectors $\boldsymbol{v}_n = (v_{n1}, \ldots, v_{nd_n})^\top$ and $\boldsymbol{w}_n = (w_{n1}, \ldots, w_{nd_n})^\top$ with finite $\ell_1$-norm, i.e., $\|\boldsymbol{v}_n\|_{\ell_1}, \|\boldsymbol{w}_n\|_{\ell_1} < \infty$, define

$$f_{0,0}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) = \sum_{\nu,\mu=1}^{d_n} v_{n\nu} w_{n\mu} a_{nj}^{(\nu)} a_{nj}^{(\mu)}, \quad f_{l,j}^{(n)}(\mathfrak{a}) = \sum_{\nu,\mu=1}^{d_n} v_{n\nu} w_{n\mu} [a_{nj}^{(\nu)} a_{n,j+l}^{(\mu)} + a_{nj}^{(\mu)} a_{n,j+l}^{(\nu)}]$$

for $j \in \{0, 1, \ldots\}$ and $l \in \{1, 2, \ldots\}$. Put $\widetilde{f}_{\ell,i}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) = \sum_{j=i}^{\infty} f_{\ell,j}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n)$, for $\ell, i \in \{0, 1, 2, \ldots\}$.

Introduce for coefficients $\mathfrak{a}$ satisfying Assumption (D) and vectors $\boldsymbol{v}_n$ and $\boldsymbol{w}_n$ the $\mathcal{F}_{nk}$-martingales

$$M_k^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) = \widetilde{f}_{0,0}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) \sum_{i=0}^{k} (\epsilon_{ni}^2 - \sigma_i^2) + \sum_{i=0}^{k} \epsilon_{ni} \sum_{\ell=1}^{\infty} \widetilde{f}_{\ell,0}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) \epsilon_{n,i-j}, \qquad k \geq 0,$$

which start in $M_0^{(n)} = 0$, for each $n \geq 0$. Put

$$S_{n',m'}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) = \sum_{i=m'+1}^{m'+n'} (Y_{ni}(\boldsymbol{v}_n^\top \boldsymbol{a}_n) Y_{ni}(\boldsymbol{w}_n^\top \boldsymbol{a}_n) - \mathrm{E}[Y_{ni}(\boldsymbol{v}_n^\top \boldsymbol{a}_n) Y_{ni}(\boldsymbol{w}_n^\top \boldsymbol{a}_n)]), \qquad m', n' \geq 0.$$

Notice that, by definitions (16) and (17),

$$S_{k,0}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n) = \boldsymbol{D}_{nk}^{(1)}, \qquad S_{k,0}^{(n)}(\mathfrak{c}, \boldsymbol{v}_n, \boldsymbol{w}_n) = \boldsymbol{D}_{nk}^{(2)}, \tag{42}$$

for $k \geq 1$ and $n \geq 1$, where $\boldsymbol{D}_{nk} = (\boldsymbol{D}_{nk}^{(1)}, \boldsymbol{D}_{nk}^{(2)})$. For brevity introduce the difference operator

$$\begin{aligned}
\delta M_{m'+n'}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) &= M_{m'+n'}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) - M_{m'}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) \\
&= \widetilde{f}_{0,0}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) \sum_{i=m'+1}^{m'+n'} (\epsilon_{ni}^2 - \sigma_{ni}^2) + \sum_{i=m'+1}^{m'+n'} \epsilon_{ni} \sum_{\ell=1}^{\infty} \widetilde{f}_{\ell,0}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) \epsilon_{n,i-\ell}, \qquad k, n \geq 1,
\end{aligned}$$

which takes the lag $n'$ forward difference at $m'$. Notice that for $m' = 0$

$$\delta M_k^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) = \widetilde{f}_{0,0}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) \sum_{i=1}^{k} (\epsilon_{ni}^2 - \sigma_{ni}^2) + \sum_{i=1}^{k} \epsilon_{ni} \sum_{\ell=1}^{\infty} \widetilde{f}_{\ell,0}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) \epsilon_{n,i-\ell}, \qquad k, n \geq 1,$$

coincides with the martingale $M_k^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n)$. A direct calculation shows that

$$\begin{aligned}
&\mathrm{Cov}(\delta M_{m'+n'}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n), \delta M_{m'+n'}^{(n)}(\mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)) \\
&= \widetilde{f}_{0,0}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n) \widetilde{f}_{0,0}^{(n)}(\mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n) \sum_{j=1}^{n'} (\gamma_{n,m'+j} + \sigma_{n,m'+j}^4) + \sum_{j=1}^{n'} \sum_{\ell=1}^{\infty} \widetilde{f}_{\ell,0}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n) \widetilde{f}_{\ell,0}^{(n)}(\mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n) \sigma_{n,m'+j}^2 \sigma_{n,m'+j-\ell}^2,
\end{aligned} \tag{43}$$

for $n', m' \geq 0$ and $n \geq 1$.

### 7.2. Martingale approximations

The following lemma provides an explicit formula for the asymptotic covariance parameter related to the two CUSUMs, $\beta_n^2(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n, \mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)$, using different pairs $(\boldsymbol{v}_n, \boldsymbol{w}_n)$ and $(\widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)$ of weighting vectors, abbreviated as $\beta_n^2(\mathfrak{b}, \mathfrak{c}) = \beta_n^2(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n, \mathfrak{c}, \boldsymbol{v}_n, \boldsymbol{w}_n)$. Especially, it follows from these results that the asymptotic variance of a single CUSUM detector under the no-change null hypothesis, $\alpha^2(\mathfrak{a}) = \beta_n^2(\mathfrak{a}, \mathfrak{a})$, satisfies

$$\alpha_n^2(\mathfrak{a}) \approx \frac{1}{n} \mathrm{Var}(D_{nn}).$$

The following general results hold under a mild condition on the error terms and especially show that (43) can be approximated by $n' \beta_n^2(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n, \mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)$ at the rate $(n')^{1-\theta}$, uniformly in $n$ and $m'$, cf. [35, (3.18)] and [21]. The proof extends these latter results and improves the bounds, but it is technical and thus deferred to the supplement. The improved bounds show that the $\ell_1$-norms of the weighting vectors may grow slowly without sacrificing the convergence of the second moments, cf. the verification of (II) and (III) in the proof of Theorem 1.

**Lemma 1.** *Let $\epsilon_{ni}$, $i \in \mathbb{Z}$, be independent with variances $\sigma_{ni}^2$ and third moments $\gamma_{ni}$ satisfying*

$$\frac{1}{n'}\sum_{i=1}^{n'} i|\sigma_{ni}^2 - s_{n1}^2| \overset{n,n'}{\ll} (n')^{-\beta}, \tag{44}$$

$$\frac{1}{n'}\sum_{i=1}^{n'} i|\gamma_{ni} - \gamma_n| \overset{n,n'}{\ll} (n')^{-\beta} \tag{45}$$

*for constants $s_{n1}^2 \in (0,\infty)$ and $\gamma_n \in \mathbb{R}$ for some $1 < \beta < 2$ with $1+\theta < \beta$. Then for $n, n' \geq 1$, with $K_n = \|\boldsymbol{v}_n\|_{\ell_1}\|\boldsymbol{w}_n\|_{\ell_1}\|\widetilde{\boldsymbol{v}}_n\|_{\ell_1}\|\widetilde{\boldsymbol{v}}_n\|_{\ell_1}$,*

$$\left|\mathrm{Cov}(M_{n'}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n), M_{n'}^{(n)}(\mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)) - (n')\beta_n^2(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n, \mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)\right| \overset{n,n'}{\ll} K_n(n')^{1-\theta}, \tag{46}$$

*and for $n, n' \geq 1$ and $m' \geq 0$*

$$\left|\mathrm{Cov}(\delta M_{m'+n'}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n), \delta M_{m'+n'}^{(n)}(\mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)) - (n')\beta_n^2(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n, \mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)\right| \overset{n,n',m'}{\ll} K_n(n')^{1-\theta}, \tag{47}$$

*if*

$$\beta_n^2(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n, \mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n) = \widetilde{f}_{0,0}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n)\widetilde{f}_{0,0}^{(n)}(\mathfrak{c})(\gamma_n - s_{n1}^4) + s_{n1}^4\sum_{\ell=1}^{\infty}\widetilde{f}_{\ell,0}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n)\widetilde{f}_{\ell,0}^{(n)}(\mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n). \tag{48}$$

**Lemma 2.** *Let $\{\epsilon_{nk} : k \geq 1, n \geq 1\}$ be independent mean zero random variables with variances $\sigma_{nk}^2$ and third moments $\gamma_{nk}$ satisfying Assumption (E). Let $\mathfrak{a}$ be coefficients satisfying the decay condition (D). Then we have for $n', m' \geq 0$ and $n \geq 1$*

$$\mathrm{E}(S_{n',m'}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n) - \delta M_{m'+n'}^{(n)}(\mathfrak{a}, \boldsymbol{v}_n, \boldsymbol{w}_n))^2 \overset{n,m',n'}{\ll} \|\boldsymbol{v}_n\|_{\ell_1}^2\|\boldsymbol{w}_n\|_{\ell_1}^2(n')^{1-\theta}. \tag{49}$$

*Further, for $k \geq 1$ and $n \geq 1$*

$$\mathrm{E}(\boldsymbol{D}_{nk}^{(1)} - \delta M_k^{(n)}(\mathfrak{b}))^2 \overset{n,k}{\ll} \|\boldsymbol{v}_n\|_{\ell_1}^2\|\boldsymbol{w}_n\|_{\ell_1}^2 k^{1-\theta}, \tag{50}$$

$$\mathrm{E}(\boldsymbol{D}_{nk}^{(2)} - \delta M_k^{(n)}(\mathfrak{c}))^2 \overset{n,k}{\ll} \|\boldsymbol{v}_n\|_{\ell_1}^2\|\boldsymbol{w}_n\|_{\ell_1}^2 k^{1-\theta}, \tag{51}$$

*such that*

$$\mathrm{E}\|\boldsymbol{D}_{nk} - \delta \boldsymbol{M}_k^{(n)}\|_2^2 \overset{n,k}{\ll} \|\boldsymbol{v}_n\|_{\ell_1}^2\|\boldsymbol{w}_n\|_{\ell_1}^2 n^{1-\theta}. \tag{52}$$

*(50)–(52) also hold (with obvious modifications), if $\boldsymbol{D}_{nk}^{(1)} = S_{k,0}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n)$ and $\boldsymbol{D}_{nk}^{(2)} = S_{k,0}^{(n)}(\mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)$ for two pairs of weighting vectors, where the bound in (52) becomes $\max\{\|\boldsymbol{v}_n\|_{\ell_1}^2\|\boldsymbol{w}_n\|_{\ell_1}^2, \|\widetilde{\boldsymbol{v}}_n\|_{\ell_1}^2\|\widetilde{\boldsymbol{w}}_n\|_{\ell_1}^2\}n^{1-\theta}$.*

**Proof.** See supplement. □

The next lemma studies the conditional covariances of the approximating martingales. It generalizes [36, Lemma 2.2] to the change-point model and two different pairs of projection vectors.

**Lemma 3.** *Suppose that the conditions of Lemma 1 hold and $\beta_n^2(\mathfrak{b}, \mathfrak{c})$ is as defined there. Then it holds for $m', n' \geq 0$ and $n \geq 1$ with $K_n = \|\boldsymbol{v}_n\|_{\ell_1}\|\boldsymbol{w}_n\|_{\ell_1}\|\widetilde{\boldsymbol{v}}_n\|_{\ell_1}\|\widetilde{\boldsymbol{w}}_n\|_{\ell_1}$*

$$E_{n'}^{(n)} = \left\|\mathrm{E}\left[(\delta M_{m'+n'}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n))(\delta M_{m'+n'}^{(n)}(\mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)) \mid \mathcal{F}_{n,m'}\right] - n'\beta_n^2(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n, \mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)\right\|_{L_1} \overset{n,m',n'}{\ll} K_n(n')^{1-\theta/2}$$

*and*

$$\left\|\mathrm{E}\left[(S_{m',n'}^{(n)}(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n))(S_{m',n'}^{(n)}(\mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)) \mid \mathcal{F}_{n,m'}\right] - n'\beta_n^2(\mathfrak{b}, \boldsymbol{v}_n, \boldsymbol{w}_n, \mathfrak{c}, \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)\right\|_{L_1} \overset{n,m',n'}{\ll} K_n(n')^{1-\theta/2}.$$

**Proof.** See supplement. □

*7.3. Proofs of Section 3.2*

After the above preparations, we are now in a position to show Theorem 1.

**Proof of Theorem 1.** Put

$$\xi_i^{(n)} = \xi_i^{(n)}(\boldsymbol{v}_n^\top \boldsymbol{b}_n, \boldsymbol{v}_n^\top \boldsymbol{c}_n) = \begin{pmatrix} Y_{ni}(\boldsymbol{v}_n^\top \boldsymbol{b}_n)Y_{ni}(\boldsymbol{w}_n^\top \boldsymbol{b}_n) - \mathrm{E}[Y_{ni}(\boldsymbol{v}_n^\top \boldsymbol{b}_n)Y_{ni}(\boldsymbol{w}_n^\top \boldsymbol{b}_n)] \\ Y_{ni}(\boldsymbol{v}_n^\top \boldsymbol{c}_n)Y_{ni}(\boldsymbol{w}_n^\top \boldsymbol{c}_n) - \mathrm{E}[Y_{ni}(\boldsymbol{v}_n^\top \boldsymbol{c}_n)Y_{ni}(\boldsymbol{w}_n^\top \boldsymbol{c}_n)] \end{pmatrix}, \tag{53}$$

such that $\boldsymbol{D}_{nk} = \sum_{i \le k} \xi_i^{(n)}$, for $k \ge 1$ and $n \ge 1$. Let us consider the bivariate extension of the sums $S_{n',m'}^{(n)}$,

$$\boldsymbol{S}_{n',m'}^{(n)} = (S_{n',m'}^{(n)}(1), S_{n',m'}^{(n)}(2))^\top = \sum_{k=m'+1}^{m'+n'} \xi_k^{(n)}, \qquad m', n' \ge 0.$$

Introduce the conditional covariance operators

$$\boldsymbol{C}_{n',m'}^{(n)}(\boldsymbol{u}) = \mathrm{E}[\boldsymbol{u}^\top S_{n',m'}^{(n)} S_{n',m'}^{(n)} | \mathcal{F}_{n,m'}], \qquad \boldsymbol{u} \in \mathbb{R}^2, \ n, n', m' \ge 1,$$

and the unconditional covariance operator associated to the Brownian motion $\boldsymbol{B}^{(n)}$,

$$\boldsymbol{T}^{(n)}(\boldsymbol{u}) = \mathrm{E}(\boldsymbol{u}^\top \boldsymbol{B}_n \boldsymbol{B}_n), \qquad \boldsymbol{u} \in \mathbb{R}^2, \ n \ge 1.$$

We shall verify [29, Th. 1], namely the validity of the following conditions: For $m' \ge 0$, $n' \ge 1$,

(I) $\sup_{j \ge 1} \mathrm{E}\|\xi_j^{(n)}\|_2^{2+\delta} < \infty$ for some $\delta > 0$.

(II) For some $\varepsilon > 0$ it holds

$$\mathrm{E}\|\mathrm{E}(\boldsymbol{S}_{n',m'}^{(n)} \mid \mathcal{F}_{n,m'})\|_2 \overset{n,n',m'}{\ll} (n')^{1/2-\varepsilon}$$

(III) There exists a covariance operator $\boldsymbol{C}$, namely $\boldsymbol{T}^{(n)}$, such that the conditional covariance operator $\boldsymbol{C}_{n',m'}^{(n)}$ converges to $\boldsymbol{C}$ in the semi-norm $\|\cdot\|_{op}$ in expectation in the sense that for some $\theta' > 0$.

$$\mathrm{E}\|\boldsymbol{C}_{n'}(\cdot \mid \mathcal{F}_{n,m'}) - \boldsymbol{C}(\cdot)\|_{op} \overset{n,n',m'}{\ll} (n')^{1-\theta'}.$$

**Remark.** As the construction is for fixed $n$, one could consider $\overset{n',m'}{\ll}$ in (II) and (III). But since we are interested in $n \to \infty$ and (II) and (III) yield the moment convergence with rate for the partial sum $\boldsymbol{S}_{n,0}^{(n)}$ of interest (for large $n$), we show $\overset{n,n',m'}{\ll}$ and consider the case $n' \ge n$. This includes the real sample size $n$ and (21) then ensures the bound $\|\boldsymbol{v}_n\|_{\ell_1}^2 \|\boldsymbol{w}_n\|_{\ell_1}^2 (n')^{-\theta/2} = O((n')^{-\theta'/2})$ we shall use.

Write $\xi_i^{(n)} = (\xi_{i1}^{(n)}, \xi_{i2}^{(n)})^\top$, $i \ge 1$, and observe that $\xi_{ij}^{(n)} = \|\boldsymbol{v}_n\|_{\ell_1} \|\boldsymbol{w}_n\|_{\ell_1} \xi_{nij}^*$ where $\xi_{nij}^*$ is obtained from $\xi_{ij}^{(n)}$ by replacing $\boldsymbol{v}_n$ by $\boldsymbol{v}_n^* = \boldsymbol{v}_n / \|\boldsymbol{v}_n\|_{\ell_1}$ and $\boldsymbol{w}_n$ by $\boldsymbol{w}_n^* = \boldsymbol{w}_n / \|\boldsymbol{w}_n\|_{\ell_1}$. The $C_r$-inequality and Cauchy–Schwarz yield

$$\begin{aligned} \mathrm{E}|\xi_{i1}^{(n)}|^{2+\delta} &\le \|\boldsymbol{v}_n\|_{\ell_1}^{2+\delta} \|\boldsymbol{w}_n\|_{\ell_1}^{2+\delta} \mathrm{E}(|Y_{ni}(\boldsymbol{v}_n^{*\top} \boldsymbol{b}_n) Y_{ni}(\boldsymbol{w}_n^{*\top} \boldsymbol{b}_n)| + \mathrm{E}|Y_{ni}(\boldsymbol{v}_n^{*\top} \boldsymbol{b}_n) Y_{ni}(\boldsymbol{w}_n^{*\top} \boldsymbol{b}_n)|)^{2+\delta} \\ &\le \|\boldsymbol{v}_n\|_{\ell_1}^{2+\delta} \|\boldsymbol{w}_n\|_{\ell_1}^{2+\delta} 2^{3+\delta} \mathrm{E}|Y_{ni}(\boldsymbol{v}_n^{*\top} \boldsymbol{b}_n) Y_{ni}(\boldsymbol{w}_n^{*\top} \boldsymbol{b}_n)|^{2+\delta} \\ &\le \|\boldsymbol{v}_n\|_{\ell_1}^{2+\delta} \|\boldsymbol{w}_n\|_{\ell_1}^{2+\delta} 2^{3+\delta} \sqrt{\mathrm{E}|Y_{ni}(\boldsymbol{v}_n^{*\top} \boldsymbol{b}_n)|^{4+2\delta}} \sqrt{\mathrm{E}|Y_{ni}(\boldsymbol{w}_n^{*\top} \boldsymbol{b}_n)|^{4+2\delta}}, \end{aligned}$$

and the second component is estimated analogously. Following the arguments in [21, p. 343], for $\delta' \in (0, 2)$ and $\chi = \delta'/2$, one can show that for $\mathfrak{a} \in \{\mathfrak{b}, \mathfrak{c}\}$ and $\boldsymbol{u}_n \in \{\boldsymbol{v}_n^*, \boldsymbol{w}_n^*\}$

$$\mathrm{E}|Y_{ni}(\boldsymbol{u}_n' \boldsymbol{a}_n)|^{4+\delta'} \le \sup_{n,k \ge 0} \mathrm{E}|\epsilon_{nk}| \sum_{\ell=0}^{\infty} |a_{n\ell}^{(u)}|^{2(2+\chi)} + \sup_{n,k \ge 0} \mathrm{E}(\epsilon_{nk}^2) \left\{ \sup_{n',k' \ge 0} \mathrm{E}(\epsilon_{n'k'}^2) \right\}^{1+\chi} \sum_{\ell=0}^{\infty} |a_{n\ell}^{(u)}|^2 \left\{ \sum_{\ell=0}^{\infty} |a_{n\ell}^{(u)}|^2 \right\}^{1+\chi},$$

where $a_{n\ell}^{(u)} = \sum_{\nu=1}^{d_n} a_{n\ell}^{(\nu)} u_{n\nu} \ll (\max(\ell, 1))^{-3/4-\theta/2}$, uniformly in uniformly $\ell_1$-bounded $\boldsymbol{u}_n$ and $n \ge 1$, such that $\sum_{\ell=0}^{\infty} |a_{n\ell}^{(u)}|^2 < \infty$ and, in turn, $\sum_{\ell=0}^{\infty} |a_{n\ell}^{(u)}|^{2(2+\chi)} < \infty$. Eventually, we obtain for any $n$

$$\max_{j=1,2} \sup_{i \ge 1} \mathrm{E}|\xi_{ij}^{(n)}|^{2+\delta} = O(\|\boldsymbol{v}_n\|_{\ell_1}^{2+\delta} \|\boldsymbol{w}_n\|_{\ell_1}^{2+\delta}). \tag{54}$$

Now Jensen's inequality yields $\mathrm{E}\|\xi_i^{(n)}\|_2^{2+\delta} = 2^{1+\delta/2} \mathrm{E}\left(\frac{1}{2}\sum_{j=1,2} [\xi_{ij}^{(n)}]^2\right)^{1+\delta/2} \le 2^{\delta/2} \sum_{j=1,2} \mathrm{E}|\xi_{ij}^{(n)}|^{2+\delta} < \infty$, verifying (I). To show (II) recall that the martingale approximation for $\boldsymbol{S}_{n',m'}^{(n)} = (S_{n',m'}^{(n)}(\mathfrak{b}), S_{n',m'}^{(n)}(\mathfrak{c}))^\top$ is given by $\delta \boldsymbol{M}_{n',m'}^{(n)} = (\delta M_{m'+n'}^{(n)}(\mathfrak{b}), \delta M_{m'+n'}^{(n)}(\mathfrak{c}))^\top$, see Lemma 2. Using

$$\mathrm{E}(\delta M_{m'+n'}^{(n)}(\mathfrak{b}) \mid \mathcal{F}_{n,m'}) = 0, \text{ and hence } \mathrm{E}(\boldsymbol{S}_{n',m'}^{(n)} \mid \mathcal{F}_{n,m'}) = \mathrm{E}(\boldsymbol{S}_{n',m'}^{(n)} - \delta \boldsymbol{M}_{n',m'}^{(n)} \mid \mathcal{F}_{n,m'}),$$

and the contraction property of conditional expectation, it follows that

$$\mathrm{E}\|\mathrm{E}(\boldsymbol{S}_{n',m'}^{(n)} \mid \mathcal{F}_{n,m'})\|_2 \overset{n,n',m'}{\ll} \|\boldsymbol{v}_n\|_{\ell_1}^2 \|\boldsymbol{w}_n\|_{\ell_1}^2 (n')^{1/2-\theta/2} \overset{n,n',m'}{\ll} (n')^{1/2-\theta'/2},$$

by virtue of Lemma 2 and (21), such that (II) holds with $\varepsilon = \theta'/2$. It remains to show (III). Observe that

$$\begin{aligned}\left\| \frac{1}{n'} \boldsymbol{C}_{n',m'}^{(n)} - \boldsymbol{T}^{(n)} \right\|_{op} &= \sup_{\boldsymbol{u}\in\mathbb{R}^2, \|\boldsymbol{u}\|_2=1} \left| \boldsymbol{u}^\top \mathrm{E}\left[ \frac{\boldsymbol{S}_{n',m'}^{(n)}}{\sqrt{n'}} \frac{\boldsymbol{S}_{n',m'}^{(n)}}{\sqrt{n'}}^\top \mid \mathcal{F}_{n,m'} \right] \boldsymbol{u} - \sum_{j=1,2} u_j \begin{pmatrix} \mathrm{Cov}(B_{n1}, B_{nj}) \\ \mathrm{Cov}(B_{n2}, B_{nj}) \end{pmatrix}^\top \boldsymbol{u} \right| \\ &= \sup_{\boldsymbol{u}\in\mathbb{R}^2, \|\boldsymbol{u}\|_2=1} \left| \sum_{i,j=1}^2 u_i u_j \left( \mathrm{E}\left[ \frac{S_{n',m'}^{(n)}(i)}{\sqrt{n'}} \frac{S_{n',m'}^{(n)}(j)}{\sqrt{n'}} \mid \mathcal{F}_{n,m'} \right] - \mathrm{Cov}(B_{ni}, B_{nj}) \right) \right|.\end{aligned}$$

Noting that $|u_i u_j| \le \max_k u_k^2 \le 1$, we obtain

$$\left\| \frac{1}{n'} \boldsymbol{C}_{n',m'}^{(n)} - \boldsymbol{T}^{(n)} \right\|_{op} \le 4 \max_{1\le i,j\le 2} \left| \mathrm{E}\left[ \frac{S_{n',m'}^{(n)}(i)}{\sqrt{n'}} \frac{S_{n',m'}^{(n)}(j)}{\sqrt{n'}} \mid \mathcal{F}_{n,m'} \right] - \mathrm{Cov}(B_{ni}, B_{nj}) \right|.$$

Therefore, (III) follows from

$$\mathrm{E}\left| \mathrm{E}\left[ \frac{S_{n',m'}^{(n)}(i)}{\sqrt{n'}} \frac{S_{n',m'}^{(n)}(j)}{\sqrt{n'}} \mid \mathcal{F}_{n,m'} \right] - \mathrm{Cov}(B_{ni}, B_{nj}) \right| \overset{n,n',m'}{\ll} \|\boldsymbol{v}_n\|_{\ell_1}^2 \|\boldsymbol{w}_n\|_{\ell_1}^2 (n')^{-\theta/2},$$

a.s., which is shown in Lemma 3, since then assumption (21) ensures the estimate $\overset{n,n',m'}{\ll} (n')^{-\theta'/2}$. Hence, from [29], we may conclude that there exists a constant $C_n$ and a universal constant $\lambda > 0$, such that

$$\|\boldsymbol{D}_{nt} - \boldsymbol{B}_n(t)\|_2 \le C_n t^{1/2-\lambda} \qquad t > 0, \tag{55}$$

a.s., which implies

$$|\boldsymbol{D}_{nt}^{(i)} - \boldsymbol{B}_n^{(i)}(t)| \le \sqrt{2} C_n t^{1/2-\lambda}, \qquad t > 0, \tag{56}$$

a.s., for $i = 1, 2$. Recalling that $D_{nt} = \boldsymbol{v}_n^\top (\boldsymbol{S}_{nt} - \mathrm{E}(\boldsymbol{S}_{nt}))\boldsymbol{w}_n$ where $\boldsymbol{S}_{nt} = \sum_{i\le t} \boldsymbol{Y}_{ni}\boldsymbol{Y}_{ni}^\top$ satisfies

$$\boldsymbol{S}_{nt} = \mathbf{1}(t \le \tau) \sum_{i\le t} \boldsymbol{Y}_{ni}(\mathfrak{b})\boldsymbol{Y}_{ni}(\mathfrak{b})^\top + \mathbf{1}(t > \tau) \left[ \sum_{i\le\tau} \boldsymbol{Y}_{ni}(\mathfrak{b})\boldsymbol{Y}_{ni}(\mathfrak{b})^\top + \sum_{i=\tau+1}^{t} \boldsymbol{Y}_{ni}(\mathfrak{c})\boldsymbol{Y}_{ni}(\mathfrak{c})^\top \right],$$

we have the following crucial representation in terms of $\boldsymbol{D}_{nt}$,

$$D_{nt} = \boldsymbol{D}_{nt}^{(1)} \mathbf{1}(t \le \tau) + [\boldsymbol{D}_{n\tau}^{(1)} + \boldsymbol{D}_{nt}^{(2)} - \boldsymbol{D}_{n\tau}^{(2)}]\mathbf{1}(t > \tau),$$

for all $t$. Since

$$\begin{aligned} &D_{nt} - \{\boldsymbol{B}_n^{(1)}(t)\mathbf{1}(t \le \tau) + [\boldsymbol{B}_n^{(1)}(\tau) + \boldsymbol{B}_n^{(2)}(t) - \boldsymbol{B}_n^{(2)}(\tau)]\} \\ &= \left(\boldsymbol{D}_{nt}^{(1)} - \boldsymbol{B}_n^{(1)}(t)\right) \mathbf{1}(t \le \tau) + \left(\boldsymbol{D}_{n\tau}^{(1)} - \boldsymbol{B}_n^{(1)}(\tau) + \boldsymbol{D}_{nt}^{(2)} - \boldsymbol{B}_n^{(2)}(t) - \boldsymbol{D}_{nt}^{(2)}(\tau) + \boldsymbol{B}_n^{(2)}(\tau)\right) \mathbf{1}(t > \tau), \end{aligned}$$

(56) yields, by definition of $G_n$, see (19),

$$|D_{nt} - G_n(t)| = |D_{nt} - \{\boldsymbol{B}_n^{(1)}(t)\mathbf{1}(t \le \tau) + [\boldsymbol{B}_n^{(1)}(\tau) + \boldsymbol{B}_n^{(2)}(t) - \boldsymbol{B}_n^{(2)}(\tau)]\}| \le 3\sqrt{2} C_n t^{1/2-\lambda},$$

for $t > 0$, a.s.. This implies

$$\frac{1}{\sqrt{n}} \max_{1\le k<n} |D_{nk} - G_n(k)| \le \frac{1}{\sqrt{n}} C_n \max_{1\le k<n} k^{1/2-\lambda} \le 3\sqrt{2} C_n n^{-\lambda}, \tag{57}$$

as $n \to \infty$, a.s., which in turn leads to (iii), since

$$\frac{1}{\sqrt{n}} \max_{1\le k<n} \left| D_{nk} - \frac{k}{n} D_{nn} - G_n^0(k) \right| = \frac{1}{\sqrt{n}} \max_{1\le k<n} \left| D_{nk} - \frac{k}{n} D_{nn} - \left[ G_n(k) - \frac{k}{n} G_n(n) \right] \right| \le 6\sqrt{2} C_n n^{-\lambda},$$

as $n \to \infty$, a.s., and (iv) follows from the reverse triangle inequality. Recalling that $U_{nk} = \mathrm{E}(U_{nk}) + D_{nk}$ and $U_{nk} - \frac{k}{n} U_{nn} = m_n(k) + D_{nk} - \frac{k}{n} D_{nn}$, we obtain

$$\frac{1}{\sqrt{n}} \max_{1\le k<n} \left| U_{nk} - \frac{k}{n} U_{nn} - \left[ m_n(k) + G_n^0(k) \right] \right| \le 6\sqrt{2} C_n n^{-\lambda},$$

as $n \to \infty$, a.s., which shows (v). (vi) now follows easily from the reverse triangle inequality. For a weight function $g$ satisfying (13) the arguments are more involved and as follows: Let $\gamma_n$ be a non-decreasing sequence specified later. Then, using $g(t)/[t(1-t)]^\beta \geq C_g$ and $n^2/(k(n-k)) \leq 2n/k$ for $1 \leq k \leq n/2$, we obtain a.s.

$$\begin{aligned}\max_{\varepsilon n/\gamma_n \leq k \leq n/2} \frac{1}{\sqrt{n} g(k/n)} \left| D_{nk} - \frac{k}{n} D_{nn} - G_n^0(k) \right| &\leq C_g^{-1} \max_{\varepsilon n/\gamma_n \leq k \leq n/2} \left( \frac{n}{k} \frac{n}{n-k} \right)^\beta \frac{1}{\sqrt{n}} \left| D_{nk} - \frac{k}{n} D_{nn} - G_n^0(k) \right| \\ &\leq C_g^{-1} (2/\varepsilon)^\beta \gamma_n^\beta \max_{1 \leq k < n} \frac{1}{\sqrt{n}} \left| D_{nk} - \frac{k}{n} D_{nn} - G_n^0(k) \right| \\ &\leq 3\sqrt{2} C_g^{-1} (2/\varepsilon)^\beta \gamma_n^\beta C_n n^{-\lambda}.\end{aligned}$$

The maximum over $n/2 \leq k \leq (1-\varepsilon/\gamma_n)n$ is estimated analogously leading to

$$\max_{\varepsilon n/\gamma_n \leq k \leq (1-\varepsilon/\gamma_n)n} \left( \frac{n}{k} \frac{n}{n-k} \right)^\beta \frac{1}{\sqrt{n}} \left| D_{nk} - \frac{k}{n} D_{nn} - G_n^0(k) \right| = O(\gamma_n^\beta C_n n^{-\lambda}), \ a.s..$$

The right-hand side is $o(1)$, a.s., if we put $\gamma_n = n^{0.1\lambda/\beta}$. Further, the technical results of the supplement and the Hájek–Rényi inequality for martingale differences yield for any $\delta > 0$ the tail bound

$$\begin{aligned}\Pr\left( \max_{1 \leq k \leq n\varepsilon/\gamma_n} \left( \frac{n}{k} \right)^\beta \frac{1}{\sqrt{n}} \left| D_{nk} - \frac{k}{n} D_{nn} \right| \geq \delta \right) &= O\left( \frac{n^{2\beta-1}}{(\delta/2)^2} \sum_{k=1}^{n\varepsilon/\gamma_n} k^{-2\beta} \right) + O\left( \frac{n^{1-\theta}}{\delta^2 n} \right) \\ &= O\left( (\varepsilon)^{1-2\beta} \gamma_n^{2\beta-1} (\log(n\varepsilon/\gamma_n) + 1)^{2\beta} \right) + o(1) \\ &= O\left( (\varepsilon)^{1-2\beta} n^{0.1(2\beta-1)\lambda/\beta} (\log(n) + 1)^{2\beta} \right) + o(1).\end{aligned}$$

The first term tends to 0, as $\varepsilon \to 0$, uniformly in $n$, since $\beta < 1/2$. Let $B^0$ be a Brownian bridge and note that $\{\alpha_n^{-1}(\mathfrak{b})\overline{G}_n^0(t) : 0 \leq t \leq \vartheta\} \stackrel{d}{=} \{B^0(t) : 0 \leq t \leq \vartheta\}$. Using the estimates $\sqrt{t}/t^\beta \leq (\varepsilon/\gamma_n)^{1/2-\beta}$ and $\log_2(1/t) \leq \log_2(n)$ on $t \in \mathcal{G}_n = \{1/n, \ldots, \lfloor \varepsilon n/\gamma_n \rfloor/n\}$ the law of the iterated logarithm for the Brownian bridge, [33, p. 72], entails for $\delta > 0$ and $\varepsilon/\gamma_n \leq \vartheta$ (thus for large $n$)

$$\Pr\left( \max_{1 \leq k \leq n\varepsilon/\gamma_n} \left( \frac{n}{k} \right)^\beta \frac{|n^{-1/2} G_n^0(k)|}{\alpha_n(\mathfrak{b})} > \delta \right) \leq \Pr\left( \sup_{t \in \mathcal{G}_n} \frac{|B^0(t)|}{\sqrt{2t \log_2(1/t)}} > \frac{\delta}{\alpha_n(\mathfrak{b}) \sqrt{2\log_2(n)} (\varepsilon/\gamma_n)^{1/2-\beta}} \right) = o(1),$$

by our choice of $\gamma_n$ and since $\beta < 1/2$. The corresponding tail probabilities for the maximum over $(1-\varepsilon/\gamma_n)n \leq k \leq n$ are treated analogously, observing the $G_n^0$ is a linear combination of Brownian motions and using the LIL for Brownian motion. Combining the above estimates shows (22). $\square$

**Proof of Theorem 2.** See supplement. $\square$

**Proof of Theorem 3.** See supplement. $\square$

**Proof of Theorem 4.** Since $\widetilde{\boldsymbol{v}}_n = a_n^{-1} \boldsymbol{v}_n$ and $\widetilde{\boldsymbol{w}}_n = b_n^{-1} \boldsymbol{w}_n$ satisfy property (21) and $D_{nk}(g; a_n^{-1}\boldsymbol{v}_n, b_n^{-1}\boldsymbol{w}_n) = a_n^{-1} b_n^{-1} D_{nk}(g; \boldsymbol{v}_n, \boldsymbol{w}_n)$, we may conclude that $T_n(g; \boldsymbol{v}_n, \boldsymbol{w}_n) = T_n(g; \widetilde{\boldsymbol{v}}_n, \widetilde{\boldsymbol{w}}_n)$. Consequently, all approximations for $T_n$ carry over. In particular, we obtain under the conditions of Theorem 1, cf. (23),

$$\left| T_n(g; \boldsymbol{v}_n, \boldsymbol{w}_n) - \max_{1 \leq k < n} \frac{1}{g(k/n)} \left| \frac{m_n(k)}{\sqrt{n}} + \overline{B}_n^0(k/n) \right| \right| = o_{\Pr}(1).$$

Note that $\overline{B}_n^0 = \alpha_n^{-1}(\mathfrak{b}) \overline{G}_n^0(t)$ is a standard Brownian bridge on $[0, \vartheta]$ and thus on the whole unit interval under the null hypothesis, whereas the scale factor changes from 1 to $\alpha_n(\mathfrak{c})/\alpha_n(\mathfrak{b})$ on $(\vartheta, 1]$. This shows (27) for $\ell_1$-bounded projections. The proof for uniformly $\ell_2$-bounded projections uses the scaling $a_n = b_n = d_n$ and the fact that by Jensen's inequality gives $\|\widetilde{\boldsymbol{v}}_n\|_{\ell_1} \leq \left( \frac{1}{d_n} \sum_{\nu=1}^\infty v_{n\nu}^2 \right)$, where the sum is finite by assumption and the factor cancels by standardization, see also [36]. $\square$

**Proof of Theorem 5.** Here is a sketch of the proof: The conditions on $g$ ensure that $\sup_{0<t<1} |B^0(t)|/g(t)$ is well defined, see [8]. Further, $\{G_n^0(t)/\alpha^2(\mathfrak{b}) : 0 \leq t \leq \tau\} \stackrel{d}{=} \{B^0(t) : 0 \leq t \leq \tau\}$ and hence a standard Brownian bridge under $H_0$. The conditions on $g$ ensure that for any (a.s.) bounded functions $h$ one has $\sup_{1/n \leq u \leq (n-1)/n} |\frac{h(u)}{g(u)} - \frac{h(u)}{g(\lfloor nu \rfloor/n)}| \to 0$ (a.s.). Further, since the drift is piecwise monotone and converges pointwise, one gets $\mu_n(\lfloor nu \rfloor/n) + \overline{B}_n^0(\lfloor nu \rfloor/n)| \to 0$ a.s. uniformly, recall Lévy's modulus of continuity, $\omega_{B^0}(a) = \sup_{0 \leq t-s \leq a} |B^0(t) - B^0(s)|$, of a Brownian bridge $B^0$, i.e. $\lim_{a \downarrow 0} \omega_{B^0}(a)/\sqrt{2a\log(1/a)} = 1$, a.s.. Using the fact that $\max_{1 \leq k < n} H_n(k/n) = \sup_{1/n \leq u/le(n-1)/n} H_n(u)$ for $H_n(u) = \mu_n(\lfloor nu \rfloor/n) + \overline{B}_n^0(\lfloor nu \rfloor/n)$ and estimating the remaining tails of the limiting process by the LIL, the result follows. (23) $\square$

**Proof of Theorem 6.** Let us stack the statistics $D_{nk}(\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj})$, as defined in (17), yielding the $2L_n$-dimensional random vector

$$\boldsymbol{D}_{nk} = \begin{pmatrix} \boldsymbol{D}_{nk}(\boldsymbol{v}_{n1}, \boldsymbol{w}_{n1}) \\ \cdots \\ \boldsymbol{D}_{nk}(\boldsymbol{v}_{nL_n}, \boldsymbol{w}_{nL_n}) \end{pmatrix} = \sum_{i \le k} \boldsymbol{\xi}_i^{(n)}, \qquad k \ge 1, \quad \boldsymbol{\xi}_i^{(n)} = \left( \boldsymbol{\xi}_{ni}^{(n)}(j) \right)_{j=1}^{L_n}.$$

Also put $\boldsymbol{S}_{n',m'}^{(n)} = \sum_{k=m'+1}^{m'+n'} \boldsymbol{\xi}_k^{(n)}$, $n', m' \ge 0$, $n \ge 1$. For sparseness of notation, we use the same symbols $\boldsymbol{D}_{nk}$, $\boldsymbol{S}_{n',m'}^{(n)}$ and $\boldsymbol{\xi}_k^{(n)}$ and note that the quantities studied here coincide with the previous definitions if $L_n = 1$. We work in the Hilbert space $\mathbb{R}^{2L_n}$ and show (I) - (III) when $L_n \to \infty$, so that the additional scaling with $L_n^{-1/2}$, which can be attached to the $\boldsymbol{\xi}_i^{(n)}$'s or put in front of the sums, is in effect. The equivalence of the vector norms $\|\cdot\|_2$ and $\|\cdot\|_\infty$ - recall that $\|\cdot\|_\infty \le \|\cdot\|_2$ and $\|\cdot\|_2 \le \sqrt{L_n}\|\cdot\|_\infty$ - and Jensen's inequality yield, in view of (54),

$$\sup_{i \ge 1} \mathrm{E}\|L_n^{-1/2}\boldsymbol{\xi}_i^{(n)}\|_2^{2+\delta} = \sup_{i \ge 1} \mathrm{E}\left[ \frac{1}{L_n} \sum_{j=1}^{L_n} \|\boldsymbol{\xi}_i^{(n)}\|_2^2 \right]^{(2+\delta)/2} \le \sup_{i \ge 1} \frac{1}{L_n} \sum_{j=1}^{L_n} \mathrm{E}\|\boldsymbol{\xi}_i^{(n)}\|_2^{2+\delta} < \infty,$$

since the bounds for $\mathrm{E}\|\boldsymbol{\xi}_i^{(n)}\|_2^{2+\delta}$ obtained above and leading to (54) are uniform in $i \ge 1$ and uniform over the considered sets of projection vectors and coefficient arrays. This shows (I). (II) follows from

$$\begin{aligned} \mathrm{E}\left\| \mathrm{E}\left( \boldsymbol{S}_{n',m'}^{(n)} \mid \mathcal{F}_{n,m'} \right) \right\|_2 &\le L_n^{1/2} \mathrm{E}\left\| \mathrm{E}\left( \boldsymbol{S}_{n',m'}^{(n)} \mid \mathcal{F}_{n,m'} \right) \right\|_\infty \\ &\le L_n^{1/2} \mathrm{E}\left( \max_{1 \le \ell \le L} \left| \mathrm{E}\left( \boldsymbol{S}_{n',m'}^{(n)}(\ell)_1 \mid \mathcal{F}_{n,m'} \right) \right| + \left| \mathrm{E}\left( \boldsymbol{S}_{n',m'}^{(n)}(\ell)_2 \mid \mathcal{F}_{n,m'} \right) \right| \right) \\ &\le L_n^{1/2} \mathrm{E}\left( \left\| \mathrm{E}\left( \boldsymbol{S}_{n',m'}^{(n)}(\cdot)_1 \mid \mathcal{F}_{n,m'} \right) \right\|_2 + \left\| \mathrm{E}\left( \boldsymbol{S}_{n',m'}^{(n)}(\cdot)_2 \mid \mathcal{F}_{n,m'} \right) \right\|_2 \right), \end{aligned}$$

such that $\mathrm{E}\left\| \mathrm{E}\left( L_n^{-1/2}\boldsymbol{S}_{n',m'}^{(n)} \mid \mathcal{F}_{n,m'} \right) \right\|_2 \overset{n,n',m'}{\ll} \|\boldsymbol{v}_n\|_{\ell_1}^2 \|\boldsymbol{w}_n\|_{\ell_1}^2 (n')^{-1/2-\theta/2} \overset{n,n',m'}{\ll} (n')^{-1/2-\theta'/2}$, by the assumptions on the growth of $\|\boldsymbol{v}_n\|_{\ell_1}^2 \|\boldsymbol{w}_n\|_{\ell_1}^2$. Next consider the conditional covariance operators

$$C_{n'm'}^{(n)}(\boldsymbol{u}) = \mathrm{E}\left( \boldsymbol{u}^\top (L_n^{-1/2}\boldsymbol{S}_{n',m'}^{(n)})(L_n^{-1/2}\boldsymbol{S}_{n',m'}^{(n)} | \mathcal{F}_{n,m'}) \right), \qquad \boldsymbol{u} \in \mathbb{R}^{2L_n},$$

and the covariance operator $\boldsymbol{T}^{(n)}(\boldsymbol{u}) = \mathrm{E}(\boldsymbol{u}^\top \boldsymbol{B}_n \boldsymbol{B}_n)$, $\boldsymbol{u} \in \mathbb{R}^{2L_n}$. We need to estimate the operator norm of their difference and use Lemma 3 and similar arguments as in the proof of Theorem 2.2 of [36]. Denote the $\nu$th coordinate of $\boldsymbol{S}_{n',m'}^{(n)}$ corresponding to the weighting vectors $\boldsymbol{v}_n(\nu)$ and $\boldsymbol{w}_n(\nu)$ by $\boldsymbol{S}_{n',m'}^{(n)}(\nu)$ and let $C_{n'm'}^{(n)}(\nu, \mu) = \mathrm{E}((L_n^{-1/2}\boldsymbol{S}_{n',m'}^{(n)}(\nu))(L_n^{-1/2}\boldsymbol{S}_{n',m'}^{(n)}(\mu) | \mathcal{F}_{n,m'}))$. By Lemma 3

$$\mathrm{E} \max_{1 \le \nu, \mu \le 2L_n} \left| C_{n'm'}^{(n)}(\nu, \mu) - \mathrm{E}(\boldsymbol{B}_n(\nu)\boldsymbol{B}_n(\mu)) \right| \ll L_n^{-1} K_n (n')^{\theta/2} \ll L_n^{-1}(n')^{-\theta'/2},$$

where $\mathrm{E}(\boldsymbol{B}_n(\nu)\boldsymbol{B}_n(\mu)) = L_n^{-1}\beta_n^2(\mathfrak{b}, \boldsymbol{v}_n(\nu), \boldsymbol{w}_n(\nu), \mathfrak{c}, \boldsymbol{v}_n(\mu), \boldsymbol{w}_n(\mu))$. Using the estimate $|\sum_{i,j} a_{ij} x_i x_j| \le L_n \|\boldsymbol{x}\|_2^2 \max_{i,j} |a_{ij}|$ for $\boldsymbol{x} = (x_1, \ldots, x_{L_n}) \in \mathbb{R}^k$ and $a_{ij} \in \mathbb{R}$, $1 \le i, j \le L_n$, we therefore obtain

$$\mathrm{E}\left\| (n')^{-1} C_{n'm'}^{(n)} - \boldsymbol{T}^{(n)} \right\|_{op} = \mathrm{E} \sup_{\boldsymbol{u} \in \mathbb{R}^{2L_n}, \|\boldsymbol{u}\|_2 = 1} \left| \boldsymbol{u}^\top \left( (n')^{-1} C_{n'm'}^{(n)} - \boldsymbol{T}^{(n)} \right) \boldsymbol{u} \right| \overset{n}{\ll} (n')^{-\theta'/2},$$

which establishes condition (III). Hence, from [29], we may conclude that there exists a constant $C_n$ and a universal constant $\lambda > 0$, such that on a new probability space for an equivalent version of $\boldsymbol{D}_{nt}$ and a Brownian motion as described in the theorem

$$\|\boldsymbol{D}_{nt} - \boldsymbol{B}_n(t)\|_2 \le C_n t^{1/2-\lambda} \qquad t > 0,$$

a.s.. The proof can now be completed along the lines of the proof of Theorem 1 with $(\boldsymbol{G}_n, \boldsymbol{G}_n^0)$ instead of $(G_n, G_n^0)$ by arguing coordinate-wise leading to

$$\left| L_n^{-1/2} C_n(\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj}) - \max_{1 \le k < n} \frac{1}{\sqrt{n}} |m_{nj}(k) - G_{nj}^0(k)| \right| \le 6\sqrt{2} C_n n^{-\lambda},$$

where the upper bound does not depend on $j$, which establishes (29). For a positive weight function a similar bound applies when considering CUSUMs taking the maximum over $\{n_0, \ldots, n_1\}$ for $n_i = \lfloor nt_i \rfloor$, $i = 1, 2$. For a weight function $g$ satisfying (13) and CUSUMs taking the maximum over $\{1, \ldots, n-1\}$ the required LIL tail bound and the martingale approximation used to apply the Hájek–Rényi inequality do not depend on $1 \le j \le L_n$ or $L_n$, such that

$$\max_{j \le L_n} \Pr\left( \left| L_n^{-1/2} C_n^g(\boldsymbol{v}_{nj}, \boldsymbol{w}_{nj}) - \max_{1 \le k < n} \frac{1}{\sqrt{n} g(k/n)} |m_{nj}(k) - G_{nj}^0(k)| \right| > \delta \right) = o(1),$$

for any $\delta > 0$. □

### 7.4. Consistency of nuisance estimators

**Proof of Theorem 7.** Fix $0 < \varepsilon < \vartheta$. We can and will assume that $n$ is large enough to ensure that $\lfloor n\varepsilon \rfloor \geq 1$ and $\lfloor n\vartheta \rfloor > h$. Denote by $\widehat{\Gamma}_n(h; d)$ the estimator $\widehat{\Gamma}_n(h)$ regarding the dimension $d$ as a formal parameter such that $\widehat{\Gamma}_n(h) = \widehat{\Gamma}_n(h; d)\big|_{d=d_n}$. In the same vain we proceed for $\widehat{\beta}_n^2$ and all other statistics arising below and write $\widehat{\beta}_n^2(d)$ etc. The assertion will then follow by showing that the consistency is uniform in the dimension $d$. By assumption $z_{ni}^{(j)} = \boldsymbol{v}_{nj}^\top \boldsymbol{Y}_{ni} \boldsymbol{w}_{nj}^\top \boldsymbol{Y}_{ni} - \mathrm{E}(\boldsymbol{v}_{nj}^\top \boldsymbol{Y}_{ni} \boldsymbol{w}_{nj}^\top \boldsymbol{Y}_{ni})$ satisfies $z_{ni}^{(j)} = \boldsymbol{v}_{nj}^\top \boldsymbol{Y}_{ni}(\mathfrak{b}) \boldsymbol{w}_{nj}^\top \boldsymbol{Y}_{ni}(\mathfrak{b}) =: z_{ni}^{(j)}(\mathfrak{b})$ for $i \leq \tau$ and $z_{ni}^{(j)} = \boldsymbol{v}_{nj}^\top \boldsymbol{Y}_{ni}(\mathfrak{c}) \boldsymbol{w}_{nj}^\top \boldsymbol{Y}_{ni}(\mathfrak{c}) =: z_{ni}^{(j)}(\mathfrak{c})$ if $i > \tau$. Put $\xi_{ni}^{(j)} = z_{ni}^{(j)} - \mathrm{E}(z_{ni}^{(j)})$ and again let $\xi_{ni}^{(j)}(\mathfrak{b}) = \xi_{ni}^{(j)}$, if $i \leq \tau$, and $\xi_{ni}^{(j)}(\mathfrak{c}) = \xi_{ni}^{(j)}$, if $\tau < i \leq n$. By Lemmas 1 and 2, $\beta_n^2(j,k) = n^{-1}\mathrm{Cov}(\boldsymbol{v}_{nj}^\top \boldsymbol{S}_{nn} \boldsymbol{w}_{nj}, \boldsymbol{v}_{nk}^\top \boldsymbol{S}_{nn} \boldsymbol{w}_{nk}) + R_n$ with $\mathrm{E}(R_n^2) = O(n^{-\theta})$. Combining this with (59), we obtain $\beta_n^2(j,k) = \sum_{h \in \mathbb{Z}} \mathrm{E}\left(\xi_{n0}^{(j)} \xi_{n,|h|}^{(k)}\right) + R_n + o(1)$. Without loss of generality we fix $(j,k) = (1,2)$ and show that $\sum_{h\in\mathbb{Z}} \widehat{\Gamma}_n(h,1,2) - \sum_{h\in\mathbb{Z}} \mathrm{E}(\xi_{n0}^{(1)} \xi_{n,|h|}^{(2)}) = o(1)$, as $n \to \infty$, where

$$\widetilde{\Gamma}_n(u;h) = \widetilde{\Gamma}_n(u;h,d) = \frac{1}{\lfloor nu \rfloor} \sum_{i=1}^{\lfloor nu \rfloor - h} \xi_{ni}^{(1)} \xi_{n,i+h}^{(2)}.$$

Here and in the sequel we omit the dependence of $\widetilde{\Gamma}_n(u;h)$ and related quantities (namely $\widetilde{\Gamma}_n(u;h,d)$ and $\Gamma(u;h,d)$ introduced below) on 1, 2, for sake of readability.

Observe that for $h \geq 0$

$$\begin{aligned}\widetilde{\Gamma}_n(u;h,d) &= \mathbf{1}(u \leq \vartheta) \frac{1}{\lfloor nu \rfloor} \sum_{i=1}^{\lfloor nu \rfloor - h} \xi_{ni}^{(1)}(\mathfrak{b}) \xi_{n,i+h}^{(2)}(\mathfrak{b}) + \mathbf{1}(u > \vartheta) \Bigg\{ \frac{\lfloor n\vartheta \rfloor - h}{\lfloor nu \rfloor} \frac{1}{\lfloor n\vartheta \rfloor - h} \sum_{i=1}^{\lfloor n\vartheta \rfloor - h} \xi_{ni}^{(1)}(\mathfrak{b}) \xi_{n,i+h}^{(2)}(\mathfrak{b}) \\ &\quad + \frac{h}{\lfloor nu \rfloor} \frac{1}{h} \sum_{i=\lfloor n\vartheta \rfloor - h+1}^{\lfloor n\vartheta \rfloor} \xi_{ni}^{(1)}(\mathfrak{b}) \xi_{n,i+h}^{(2)}(\mathfrak{c}) + \frac{\lfloor nu \rfloor - \lfloor n\vartheta \rfloor - h}{\lfloor nu \rfloor} \frac{1}{n - \lfloor n\vartheta \rfloor - h} \sum_{i=\lfloor n\vartheta \rfloor + 1}^{\lfloor nu \rfloor - h} \xi_{ni}^{(1)}(\mathfrak{c}) \xi_{n,i+h}^{(2)}(\mathfrak{c}) \Bigg\}.\end{aligned}$$

Define for $|h| \leq m_n$ and $\mathfrak{a} \in \{\mathfrak{b}, \mathfrak{c}\}$

$$\Gamma(h,d,\mathfrak{a}) = \mathrm{E}\left(\xi_{ni}^{(1)}(\mathfrak{a}) \xi_{n,i+|h|}^{(2)}(\mathfrak{a})\right).$$

Then for $0 \leq h \leq m_n$,

$$\begin{aligned}\mathrm{E}(\widetilde{\Gamma}_n(u;h,d)) &= \mathbf{1}(u \leq \vartheta) \frac{\lfloor nu \rfloor - h}{\lfloor nu \rfloor} \Gamma(h,d,\mathfrak{b}) + \mathbf{1}(u > \vartheta) \Bigg( \frac{\lfloor n\vartheta \rfloor - h}{\lfloor nu \rfloor} \Gamma(h,d,\mathfrak{b}) + \frac{h}{\lfloor nu \rfloor} \mathrm{E}\left(\xi_{ni}^{(1)}(\mathfrak{b}) \xi_{n,i+h}^{(2)}(\mathfrak{c})\right) \\ &\quad + \frac{\lfloor nu \rfloor - \lfloor n\vartheta \rfloor - h}{\lfloor nu \rfloor} \Gamma(h,d,\mathfrak{c}) \Bigg).\end{aligned}$$

Using $|h| \leq m_n = o(n)$ and $|\frac{\lfloor na \rfloor}{\lfloor nb \rfloor} - a/b| = O(b|\lfloor na \rfloor/n - a| + a|\lfloor nb \rfloor/n - b|) = O(1/n) = o(m_n^{-1})$ for $0 < \varepsilon \leq a, b$, uniformly in $a, b$, we obtain

$$\mathrm{E}(\widetilde{\Gamma}_n(u;h)) = \mathbf{1}(u \leq \vartheta)\Gamma(h,\mathfrak{b}) + \mathbf{1}(u > \vartheta)((\vartheta/u)\Gamma(h,\mathfrak{b}) + (1 - \vartheta/u)\Gamma(h,\mathfrak{c}) + o(m_n^{-1})),$$

as $n \to \infty$, for $|h| \leq m_n$, where the $o(1)$ term is uniform in $|h| \leq m_n$ and $u \in [\varepsilon, 1]$. Consequently,

$$\begin{aligned}\sum_{|h| \leq m_n} w_{mh} \mathrm{E}(\widetilde{\Gamma}_n(u;h,d)) &= \mathbf{1}(u \leq \vartheta) \sum_{|h| \leq m_n} w_{mh} \Gamma(h,d,\mathfrak{b}) \\ &\quad + \mathbf{1}(u > \vartheta) \left( (\vartheta/u) \sum_{|h| \leq m_n} w_{mh} \Gamma(h,d,\mathfrak{b}) + (1 - \vartheta/u) \sum_{|h| \leq m_n} w_{mh} \Gamma(h,d,\mathfrak{c}) \right) + o(1),\end{aligned}$$

as $n \to \infty$, where the $o(1)$ term is uniform in $d \in \mathbb{N}$ and $u \in [\varepsilon, 1]$, such that

$$\sup_{u \in [\varepsilon,1]} \sup_{d \in \mathbb{N}} \max_{|h| \leq m_n} \left| \sum_{|h| \leq m_n} w_{mh} \mathrm{E}(\widetilde{\Gamma}_n(u;h,d)) - \sum_{|h| \leq m_n} w_{mh} \Gamma(u;h,d) \right| = o(1), \tag{58}$$

as $n \to \infty$, where

$$\Gamma(u;h,d) = \mathbf{1}(u \leq \vartheta)\Gamma(h,d,\mathfrak{b}) + \mathbf{1}(u > \vartheta)((\vartheta/u)\Gamma(h,d,\mathfrak{b}) + (1 - \vartheta/u)\Gamma(h,d,\mathfrak{c}))$$

for $u \in [\varepsilon, 1]$. As in [35, Th. 4.4] one can show that

$$\sum_{h \in \mathbb{Z}} \sup_{d \in \mathbb{N}} \left| \mathrm{E}\left(\xi_1^{(1)}(\mathfrak{a}) \xi_{1+h}^{(2)}(\mathfrak{a})\right) \right| < \infty \tag{59}$$

for $\mathfrak{a} \in \{\mathfrak{b}, \mathfrak{c}\}$ as well as $\beta^2(\mathfrak{b}) = \sum_{h\in\mathbb{Z}} \Gamma(h, d, \mathfrak{b})$ and $\beta^2(\mathfrak{c}) = \sum_{h\in\mathbb{Z}} \Gamma(h, d, \mathfrak{c})$. This implies

$$\sup_{u\in[\varepsilon,1]} \sup_{d\geq 1} \sum_{h\in\mathbb{Z}} |\Gamma(u; h, d)| < \infty, \tag{60}$$

since $\sum_{h\in\mathbb{Z}} |\Gamma(u; h, d)| \leq 2\sum_{h\in\mathbb{Z}} |\Gamma(h, d, \mathfrak{b})| + \sum_{h\in\mathbb{Z}} |\Gamma(h, d, \mathfrak{c})|$. Therefore, we may further conclude that

$$\begin{aligned}\beta^2(u; d) &:= \sum_{h\in\mathbb{Z}} \mathrm{E}(\widetilde{\Gamma}_n(u; h, d)) = \mathbf{1}(u \leq \vartheta) \sum_{h\in\mathbb{Z}} \Gamma(h, d, \mathfrak{b}) \\ &\quad + \mathbf{1}(u > \vartheta)\left( (\vartheta/u) \sum_{h\in\mathbb{Z}} \Gamma(h, d, \mathfrak{b}) + (1 - \vartheta/u) \sum_{h\in\mathbb{Z}} \Gamma(h, d, \mathfrak{c}) \right) + o(1),\end{aligned}$$

yielding the representation

$$\beta_n^2(u; d) = \sum_{h\in\mathbb{Z}} \Gamma(u; h, d) + o(1) \tag{61}$$

as well as

$$\beta_n^2(u; d) = \beta^2(u; d) + o(1), \tag{62}$$

as $n \to \infty$, uniformly in $d \in \mathbb{N}$ and $u \in [\varepsilon, 1]$, where

$$\beta^2(u; d) = \sigma^2(u; d, \mathfrak{b}, \mathfrak{c}) = \mathbf{1}(u \leq \vartheta)\beta^2(\mathfrak{b}) + \mathbf{1}(u > \vartheta)((\vartheta/u)\beta^2(\mathfrak{b}) + (1 - \vartheta/u)\beta^2(\mathfrak{c})).$$

The arguments used in the proof of [35, Th. 4.4] to obtain (A.11) therein show that, if applied to the subseries $\{\xi_{ni}^{(j)} : 1 \leq i \leq \lfloor n\vartheta \rfloor\}$ and $\{\xi_{ni}^{(j)} : \lfloor n\vartheta \rfloor + 1 \leq i \leq n - h\}$,

$$\left\| \sum_{i=1}^{\lfloor nu \rfloor} \xi_{ni}^{(j)}(\mathfrak{b}) \right\|_{L_2}^2 = C_1 \lfloor nu \rfloor, u \leq \vartheta, \quad \text{and} \quad \left\| \sum_{i=\lfloor n\vartheta \rfloor+1}^{\lfloor nu \rfloor - h} \xi_{ni}^{(j)}(\mathfrak{c}) \right\|_{L_2}^2 = C_2\left(\lfloor nu \rfloor - \lfloor n\vartheta \rfloor\right), u > \vartheta, \tag{63}$$

for constants $C_1, C_2 < \infty$ not depending on $h$, $j = 1, 2$. Hence

$$\left\| \sum_{i=1}^{\lfloor nu \rfloor - h} \xi_{ni}^{(j)} \right\|_{L_2} \leq C_3 \left( \max(\sqrt{\lfloor nu \rfloor}, \sqrt{\lfloor n\vartheta \rfloor} + \sqrt{\lfloor nu \rfloor - \lfloor n\vartheta \rfloor - h}) \right), \tag{64}$$

for $j = 1, 2$, and in turn

$$\sup_{u\in(\varepsilon,1]} \sup_{d\in\mathbb{N}} \max_{|h|\leq m_n} \|\widetilde{\Gamma}_n(u; h, d) - \mathrm{E}(\widetilde{\Gamma}_n(u; h, d))\|_{L_1} \leq C_4 n^{-1/2}, \tag{65}$$

for constants $C_3, C_4 < \infty$. Now observe that $\widehat{\Gamma}_n(u; h, d) = \frac{1}{\lfloor nu \rfloor} \sum_{i=1}^{\lfloor nu \rfloor - h} (\xi_{ni}^{(1)} - \overline{\xi}_n^{(1)})(\xi_{n,i+h}^{(2)} - \overline{\xi}_n^{(2)})$ where $\overline{\xi}_n^{(j)}(u) = \lfloor nu \rfloor^{-1} \sum_{i=1}^{\lfloor nu \rfloor} \xi_{ni}^{(j)}$, $j = 1, 2$. It holds

$$\lfloor nu \rfloor \left( \widehat{\Gamma}_n(u; h, d) - \widetilde{\Gamma}_n(u; h, d) \right) = -\overline{\xi}_n^{(1)}(u) \sum_{j=1}^{\lfloor nu \rfloor - h} \xi_{n,j+h}^{(2)} - \overline{\xi}_n^{(2)}(u) \sum_{j=1}^{\lfloor nu \rfloor - h} \xi_{nj}^{(1)} - \overline{\xi}_n^{(1)}(u) \sum_{j=1}^{\lfloor nu \rfloor} \xi_{nj}^{(2)}.$$

Again decomposing the sums as $\sum_{j=1}^{\lfloor nu \rfloor - h} = \mathbf{1}(u \leq \vartheta) \sum_{j=1}^{\lfloor nu \rfloor - h} + \mathbf{1}(u > \vartheta) \left\{ \sum_{j=1}^{\lfloor n\vartheta \rfloor - h} + \sum_{j=\lfloor n\vartheta \rfloor - h+1}^{\lfloor n\vartheta \rfloor} + \sum_{j=\lfloor n\vartheta \rfloor + 1}^{\lfloor nu \rfloor - h} \right\}$ and using (63), we obtain $\mathrm{E}\left( n|\widehat{\Gamma}_n(u; h, d) - \widetilde{\Gamma}_n(u; h, d)| \right) = O(1)$, uniformly over $|h| \leq m_n$, $d \in \mathbb{N}$ and $u \in [\varepsilon, 1]$. For example, for $0 \leq h \leq m_n$

$$\mathrm{E}\left| \overline{\xi}_n^{(2)}(u) \sum_{j=1}^{\lfloor nu \rfloor - h} \xi_{nj}^{(1)} \right| \leq \frac{1}{\lfloor nu \rfloor} \mathrm{E}\left| \sum_{i=1}^{\lfloor nu \rfloor} \xi_{ni}^{(2)} \sum_{j=1}^{\lfloor nu \rfloor - h} \xi_{nj}^{(1)} \right| \leq C_1 C_2 \left( \frac{\sqrt{\lfloor nu \rfloor}\sqrt{\lfloor nu \rfloor - h}}{\lfloor nu \rfloor} \right) = O(1),$$

We may conclude that $\sup_{d\in\mathbb{N}} \sup_{u\in[\varepsilon,1]} m_n \max_{|h|\leq m_n} \mathrm{E}|\widehat{\Gamma}_n(u; h, d) - \widetilde{\Gamma}_n(u; h, d)| = O(m_n/n) = o(1)$, as $n \to \infty$, and by boundedness of the weights it follows that

$$\sup_{d\in\mathbb{N}} \mathrm{E}\left| \sum_{|h|\leq m_n} w_{mh} \widehat{\Gamma}_n(h; d) - \sum_{|h|\leq m_n} w_{mh} \widetilde{\Gamma}_n(h; d) \right| = o(1).$$

Now, having in mind (61) and (62), decompose

$$\sum_{|h|\le m_n} w_{mh}\widetilde{\Gamma}_n(u;h,d) - \alpha^2(u,\mathfrak{b},\mathfrak{c}) = \sum_{|h|\le m_n} w_{mh}\left[\widetilde{\Gamma}_n(u;h,d) - \Gamma(u;h,d)\right] - \sum_{|h|>m_n} w_{mh}\Gamma(u;h,d),$$

and combine (58), (60) and (65), see the supplement for details. □

*7.5. Consistency of the change-point estimators*

**Proof of Theorem 9.** Observe that, by the definitions of $U_{nk}$, $D_{nk}$ and $\widetilde{D}_{nk}$,

$$\widehat{\mathcal{U}}_n(k) - \mathcal{U}_n(k) = \frac{1}{g(k/n)n}\left(\widetilde{D}_{nk} - \frac{k}{n}\widetilde{D}_{nn}\right) + \frac{R_{nk}}{g(k/n)},$$

with remainder $R_{nk} = \frac{1}{n}\left(D_{nk} - \frac{k}{n}D_{nn} - [\widetilde{D}_{nk} - \frac{k}{n}\widetilde{D}_{nn}]\right)$. By Eq. (21) in the supplement, we have for $k \le n/2$

$$\max_{1\le k\le n/2} E\left(\frac{R_{nk}}{g(k/n)}\right)^2 \le 2 \max_{1\le k\le n/2} E\left(\left(\frac{n}{k}\right)^{\beta} R_{nk}\right)^2 \overset{n}{\ll} n^{-1-\theta}$$

and the same bound holds for $n/2 < k < n$. Therefore for any $\delta > 0$

$$\Pr\left(\max_{1\le k<n}\frac{|R_{nk}|}{g(k/n)} > \delta\right) \le \Pr\left(\sum_{k=1}^{n}\left(\frac{R_{nk}}{g(k/n)}\right)^2 > \delta^2\right) \ll n^{-\theta}.$$

Hence, it suffices to show that for all $\delta > 0$ $\Pr\left(|\widetilde{D}_{nn}| > \delta n\right) = o(1)$ and $\Pr\left(\max_{1\le k<n}|\widetilde{D}_{nk}| > \delta n\right) = o(1)$, where the first assertion follows from the latter maximal inequality. Of course $E(\widetilde{D}_{nn}^2) = O(n)$, since $\widetilde{D}_{nn}$ is the sum of $n$ martingale differences. Now an application of Doob's maximal inequality entails $\Pr\left(\max_{1\le k<n}|\widetilde{D}_{nk}|^2 > \delta^2 n^2\right) = \frac{E(\widetilde{D}_{nn}^2)}{\delta^2 n^2} = O\left(\frac{1}{n}\right)$, which establishes $\Pr\left(\max_{1\le k<n}|\widetilde{D}_{nk}| > \delta n\right) = o(1)$ and in turn (40). Next consider

$$\sup_{t\in[0,1]}|\widehat{u}_n(t) - u(t)| \le \max_{1\le k<n}|\widehat{\mathcal{U}}_n(k) - \mathcal{U}_n(k)| + \sup_{t\in[0,1]}|u_n(t) - u(t)| = \sup_{t\in[0,1]}|u_n(t) - u(t)| + o_{\Pr}(1),$$

as $n \to \infty$, by (40). Clearly, $u_n(t) \to u(t)$ for each fixed $t$, and by monotonicity on $[0,\vartheta]$ and $[\vartheta, 1]$ this implies uniform convergence, since $g$ is continuous, which completes the proof. □

**Proof of Theorem 10.** Since $\vartheta \in (0, 1)$ is an isolated maximum of $u$ and $\widehat{u}_n$ converges uniformly to $u$, the consistency follows from well known results, see, e.g., [37], by virtue of Theorem 9 and (39). □

## CRediT authorship contribution statement

**Ansgar Steland:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

## Acknowledgments

The author acknowledges support from Deutsche Forschungsgemeinschaft (grants STE 1034/11-1, 1034/11-2).

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2019.104582.

## References

[1] A. Aue, S. Hörmann, L. Horváth, M. Reimherr, Break detection in the covariance structure of multivariate time series models, Ann. Statist. 37 (6B) (2009) 4046–4087.
[2] V. Avanesov, N. Buzun, Change-point detection in high-dimensional covariance structure, 2018, arXiv:1610.03783.
[3] I. Berkes, E. Gombay, L. Horváth, Testing for changes in the covariance structure of linear processes, J. Statist. Plann. Inference 139 (6) (2009) 2044–2063.
[4] A. Birnbaum, I.M. Johnstone, B. Nadler, D. Paul, Minimax bounds for sparse PCA with noisy high-dimensional data, Ann. Statist. 41 (3) (2013) 1055–1084.
[5] J. Breitung, S. Eickmeier, Testing for structural breaks in dynamic factor models, J. Econometrics 163 (1) (2011) 71–84.
[6] T. Cai, Z. Ma, Y. Wu, Optimal estimation and rank detection for sparse spiked covariance matrices, Probab. Theory Related Fields 161 (3–4) (2015) 781–815.
[7] H. Cho, P. Fryzlewicz, Multiple-change-point detection for high dimensional time series via sparsified binary segmentation, J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (2) (2015) 475–507.
[8] M. Csörgő, L. Horváth, Weighted Approximations in Probability and Statistics, in: Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Ltd., Chichester, 1993, p. xvi+442, With a foreword by David Kendall.

[9] M. Csörgő, L. Horváth, Limit Theorems in Change-Point Analysis, in: Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 1997, p. xvi+414, With a foreword by David Kendall.
[10] C. Davis, W.M. Kahan, The rotation of eigenvectors by a perturbation. III, SIAM J. Numer. Anal. 7 (1970) 1–46.
[11] N.B. Erichson, P. Zheng, K. Manohar, S.L. Brunton, J.N. Kutz, A.A. Aravkin, Sparse principal component analysis via variable projection, 2018, arXiv:1804.00341.
[12] D. Ferger, On the supremum of a Brownian bridge standardized by its maximizing point with applications to statistics, Statist. Probab. Lett. 134 (2018) 63–69.
[13] P. Galeano, D. Peña, Covariance changes detection in multivariate time series, J. Statist. Plann. Inference 137 (1) (2007) 194–211.
[14] X. Han, A. Inoue, Tests for parameter instability in dynamic factor models, Econometric Theory 31 (5) (2015) 1117–1152.
[15] L. Horváth, G. Rice, Asymptotics for empirical eigenvalue processes in high-dimensional linear factor models, J. Multivariate Anal. 169 (2019) 138–165.
[16] W.B. Johnson, J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, in: Conference in Modern Analysis and Probability (New Haven, Conn., 1982), in: Contemp. Math, vol. 26, Amer. Math. Soc., Providence, RI, 1984, pp. 189–206.
[17] I.M. Johnstone, A.Y. Lu, On consistency and sparsity for principal components analysis in high dimensions, J. Amer. Statist. Assoc. 104 (486) (2009) 682–693.
[18] L. Jolliffe, N. Trendafilov, M. Uddin, A modified principal component technique based on the lasso, J. Comput. Graph. Statist. 12 (2003) 531–547.
[19] S. Jung, J.S. Marron, PCA consistency in high dimension, low sample size context, Ann. Statist. 37 (6B) (2009) 4104–4130.
[20] D.M. Kane, J. Nelson, Sparser Johnson-Lindenstrauss transforms, J. ACM 61 (1) (2014) Art. 4, 23.
[21] M.A. Kouritzin, Strong approximation for cross-covariances of linear variables with long-range dependence, Stochastic Process. Appl. 60 (2) (1995) 343–353.
[22] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Anal. 88 (2) (2004) 365–411.
[23] X. Li, Z. Zhao, Testing for changes in autocovariances of nonparametric time series models, J. Statist. Plann. Inference 143 (2) (2013) 237–250.
[24] Z. Ma, Sparse principal component analysis and iterative thresholding, Ann. Statist. 41 (2) (2013) 772–801.
[25] G. Marsaglia, W. Wan, J. Wang, Evaluating Kolmogorov's distribution, J. Stat. Softw. 8 (18) (2003).
[26] N. Mause, A. Steland, Detecting changes in the second moment structure of high-dimensional sensor-type data in a $K$-sample setting, Sequential Anal. (2020) in press.
[27] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, Statist. Sinica 17 (4) (2007) 1617–1642.
[28] D. Paul, I.M. Johnstone, Augmented sparse principal component analysis for high dimensional data, Technical Report, 2007, arXiv:1202.1242, arXiv:1202.1242.
[29] W. Philipp, A note on the almost sure approximation of weakly dependent random variables, Monatsh. Math. 102 (3) (1986) 227–236.
[30] A. Sancetta, Sample covariance shrinkage for high dimensional dependent data, J. Multivariate Anal. 99 (5) (2008) 949–967.
[31] M. Schweinberger, S. Babkin, K.B. Ensor, High-dimensional multivariate time series with additional structure, J. Comput. Graph. Statist. 26 (3) (2017) 610–622.
[32] D. Shen, H. Shen, J.S. Marron, Consistency of sparse PCA in high dimension, low sample size contexts, J. Multivariate Anal. 115 (2013) 317–333.
[33] G.R. Shorack, J.A. Wellner, Empirical Processes with Applications to Statistics, in: Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986.
[34] A. Steland, Random walks with drift—a sequential approach, J. Time Series Anal. 26 (6) (2005) 917–942.
[35] A. Steland, R. von Sachs, Large-sample approximations for variance-covariance matrices of high-dimensional time series, Bernoulli 23 (4A) (2017) 2299–2329.
[36] A. Steland, R. von Sachs, Asymptotics for high-dimensional covariance matrices and quadratic forms with applications to the trace functional and shrinkage, Stochastic Process. Appl. 128 (8) (2018) 2816–2855.
[37] A.W. van der Vaart, Asymptotic Statistics, Cambridge University Press, Cambridge, 1998, p. xvi+443.
[38] L.J. Vostrikova, Discovery of disorder in multidimensional random processes, Dokl. Akad. Nauk SSSR 259 (2) (1981) 270–274.
[39] W. Wang, J. Fan, Asymptotics of empirical eigenstructure for high dimensional spiked covariance, Ann. Statist. 45 (3) (2017) 1342–1374.
[40] T. Wang, R.J. Samworth, High dimensional change point estimation via sparse projection, J. R. Stat. Soc. Ser. B Stat. Methodol. 80 (1) (2018) 57–83.
[41] Y. Yu, T. Wang, R.J. Samworth, A useful variant of the Davis-Kahan theorem for statisticians, Biometrika 102 (2) (2015) 315–323.