

TESTING AND ESTIMATING CHANGE-POINTS IN THE COVARIANCE MATRIX OF A HIGH-DIMENSIONAL TIME SERIES

ANSGAR STELAND

ABSTRACT. 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.

Keywords: Big data, Change-point, CUSUM transform, Data science, High-dimensional statistics, Projection, Spatial statistics, Spiked covariance, Strong approximation, VARMA processes
MSC: 62E20, 62M10, 62H99

1. INTRODUCTION

High-dimensional big data arise in diverse fields such as environmetrics, engineering and finance. From a data science viewpoint statistical methods and tools are needed, which allow to answer questions posed to the data, and mathematical results justifying their validity under mild regularity conditions. The latter especially requires asymptotics for the case that the data dimension is large in comparison to the sample size. In this paper, a high-dimensional time series model is studied and all asymptotic results allow for increasing dimension without any constraint relative to the sample size. The proposed procedures are investigated by simulations and applied to real data from environmetrics.

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 ℓ_1 - resp. ℓ_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 approach, [Cho and Fryzlewicz, 2015] 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, [Galeano and Peña, 2007], 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. [Berkes et al., 2009] 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 [Steland, 2005] and [Li and Zhao, 2013]. [Aue et al., 2009] 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 $d/n \rightarrow y \in (0, 1)$, as typically imposed in random matrix theory, $d = O(h(n))$ for some increasing function h , e.g., exponential growth as in [Avanesov and Buzun, 2018] (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 λ_j , [Wang and Fan, 2017], which allows for $d/n \rightarrow \infty$ 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 [Steland and von Sachs, 2017] and one can also consider an increasing number of such statistics by virtue of the results in [Steland and von Sachs, 2018]. 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 [Steland and von Sachs, 2017, Steland and von Sachs, 2018], 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 [Ledoit and Wolf, 2004] and [Sancetta, 2008], 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 [Steland and von Sachs, 2017] and [Steland and von Sachs, 2018]. 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 that 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 [Vostrikova, 1981], 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 A, whereas additional material is deferred to an appendix.

2. MODEL, ASSUMPTIONS AND PROCEDURES

2.1. Notation. Throughout the paper $a_{nk} \stackrel{n,k}{\ll} b_{nk}$ for two arrays of real numbers means that there exists a constant $C < \infty$, such that $a_{nk} \leq C b_{nk}$ for all n, k . $(\Omega, \mathcal{A}, \mathbb{P})$ denotes the underlying probability space on which the vector time series is defined. \mathbb{E} denotes expectation (w.r.t. \mathbb{P}), \mathbb{V} the variance and \mathbb{C} the covariance. 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 ℓ_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 $\mathbf{Y}_{ni} = (Y_{ni}^{(1)}, \dots, Y_{ni}^{(d_n)})^\top$ are given by

$$(1) \quad Y_{ni}^{(\nu)} = Y_{ni}^{(\nu)}(\mathbf{a}) = \sum_{j=0}^{\infty} a_{nj}^{(\nu)} \epsilon_{n,i-j}, \quad i \in \{1, \dots, n\}, \nu \in \{1, \dots, d_n\}, n \geq 1,$$

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

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

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

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

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

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

The assumptions on σ_{ni}^2 and γ_{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 $\sum_{i=1}^{\ell} |\kappa_i| \sim \ell^{1+\beta}$.

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 $\lambda_1 > \dots > \lambda_r > 0$ and let $\mathbf{u}_{nj} = (u_{nj}^{(\nu)})_{\nu=1}^d \in \mathbb{R}^d$, $j \in \{1, \dots, r\}$, be orthonormal vectors with $\|\mathbf{u}_{nj}\|_{\ell_1} \leq C$ for $j \in \{1, \dots, r\}$. Assume that

$$(3) \quad \Sigma_n = \sum_{j=1}^r \lambda_j \mathbf{u}_{nj} \mathbf{u}_{nj}^\top + \sigma^2 I_{d_n}.$$

The r leading eigenvalues of Σ_n under model (3) are $\lambda_j + \sigma^2$, $j \in \{1, \dots, r\}$, and represent spikes in the spectrum, which is otherwise flat and given by σ^2 . The assumption that the eigenvectors are ℓ_1 -bounded is common in high-dimensional statistics, especially when assuming a

spiked covariance model: [Johnstone and Lu, 2009] have shown that principal component analysis (PCA) generates inconsistent estimates of the leading eigenvectors if $d/n \rightarrow y \in (0, 1)$, which motivated developments on sparse PCA. Minimax bounds for sparse PCA have been studied by [Birnbbaum et al., 2013] under ℓ_q -constraints on the eigenvectors for $0 < q < 2$. For example, the simple diagonal thresholding estimator $\hat{\mathbf{u}}_{nj}^{th}$ of the j th leading eigenvector \mathbf{u}_{nj} of [Johnstone and Lu, 2009] satisfies $\mathbb{E}\|\hat{\mathbf{u}}_{nj}^{th} - \hat{\mathbf{u}}_{nj}^{th\top} \mathbf{u}_{nj} \mathbf{u}_{nj}\|_2^2 = O(n^{-1/4})$ and the iterated version of [Ma, 2013] attains the optimal rate $O(n^{(1-q/2)})$, see also [Paul and Johnstone, 2007]. An ℓ_1 sparseness assumption on the eigenvectors is weaker than the (joint) k -sparseness condition on the row support of matrix of eigenvectors imposed in [Cai et al., 2015], who study optimal estimation under the spectral norm.

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

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 $0 = r_1 \leq r_2 \leq \dots \leq r_q$ be integers and define the q -variate innovations

$$\boldsymbol{\epsilon}_{ni} = (\epsilon_{n,i-r_1}, \dots, \epsilon_{n,i-r_q})^\top, \quad i \geq 1, n \geq 1,$$

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

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

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

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

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

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

then the coefficients $c_{nk}^{(Z,\nu)} = \sum_{\ell=1}^q \mathbf{1}(k \geq r_\ell) b_{n,k-r_\ell}^{(\nu,\ell)}$ of the series (4) satisfy $\sup_{n \geq 1} \max_{\nu \geq 1} |c_{nk}^{(Z,\nu)}| \ll k^{-3/4-\theta/2}$, i.e., Assumption (D) holds. In this construction the lags r_1, \dots, 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 \mathbf{e}_{ni} , $i \geq 1$, with a covariance matrix close to some $\mathbf{V} > 0$: Let

$$(5) \quad \mathbf{Z}_{ni} = \sum_{j=0}^{\infty} \mathbf{B}_{nj} \mathbf{P} \mathbf{e}_{n,i-j}, \quad \mathbf{e}_{ni} = \mathbf{V}^{1/2} \boldsymbol{\epsilon}_{ni}, \quad i \geq 1, n \in \mathbb{N},$$

where

$$(6) \quad \epsilon_{ni} = (\epsilon_{n,i-r_1}, \dots, \epsilon_{n,i-r_{d_n}})^\top, \quad i \geq 1, n \in \mathbb{N},$$

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

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

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

with $\varpi = 0$ and assuming that the operators $\mathbf{B}_{nj}\mathbf{P}\mathbf{V}^{1/2}$, $n \geq 1$, are trace class operators in the sense that $\sum_i |\pi_{ni}| = O(1)$, with eigenvectors satisfying $\|\mathbf{l}_{ni}\|_{\ell_1}, \|\mathbf{r}_{ni}\|_{\ell_1} \stackrel{n,i}{\ll} 1$. For $q = d_n$ we let $\mathbf{P} = \mathbf{I}$ such that $\mathbf{l}_{ni} = \mathbf{r}_{ni}$ are the eigenvectors and π_{ni} the eigenvalues of \mathbf{V} . Then $\sup_{n \geq 1} \max_{\nu \geq 1} |c_{nj}^{(Z, \nu)}| \ll j^{-3/4-\theta/2} \sum_{\ell=1}^{\infty} r_\ell^{-3/4-\theta/2} \ll j^{-3/4-\theta/2}$ verifying (D), as shown in the appendix. The ℓ_1 constraint on the eigenvectors can be omitted when imposing the stronger condition $\sup_{\nu \geq 1} |\sum_{j=1}^{d_n} b_{n,k-r_\ell}^{(\nu, j)}|^2 \stackrel{n}{\ll} (k + r_\ell)^{-3-\theta}$ on the coefficient matrices. For details see the appendix.

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

$$\mathbf{Y}_{ni} = \mathbf{A}_{n1}\mathbf{Y}_{n,i-1} + \dots + \mathbf{A}_{np}\mathbf{Y}_{n,i-p} + \mathbf{M}_{n1}\epsilon_{n,i-1} + \dots + \mathbf{M}_{nr}\epsilon_{n,i-r} + \epsilon_{ni},$$

with colored d_n -variate innovations as in (A.1). $\mathbf{A}_{n1}, \dots, \mathbf{A}_{np}$ and $\mathbf{M}_{n1}, \dots, \mathbf{M}_{nr}$ are $(d_n \times d_n)$ coefficient matrices. Let us assume that each of these coefficient matrices satisfies (7) with $\varpi = 1$ for some $\delta > 0$, when denoting its elements by $b_{nj}^{(\nu, \ell)}$, $1 \leq \nu, \ell \leq d_n$. Recall that the process is stable, if $\det(I_{d_n} - \mathbf{A}_{n1}z - \dots - \mathbf{A}_{np}z^p) \neq 0$ for $|z| \leq 1$. Then the operator $\mathbf{A}(L) = I_{d_n} - \sum_{j=1}^p \mathbf{A}_{nj}L^j$, where L denotes the lag operator, is invertible, the coefficient matrices, \mathbf{D}_{nj} , of $\Psi(L) = \mathbf{A}(L)^{-1}$ are absolutely summable, and one obtains the MA representation $\mathbf{Y}_{ni} = \left(\sum_{j=0}^{\infty} \mathbf{D}_{nj}L^j\right) \left(\sum_{k=1}^r \mathbf{M}_{nk}L^k\right) \epsilon_{n,i-j} = \sum_{j=0}^{\infty} \Phi_{nj} \epsilon_{n,i-j}$. As well known, the coefficient matrices, Φ_{nj} , can be calculated using the recursion $\Phi_{n0} = I_{d_n}$, $\Phi_{nj} = \mathbf{M}_{nj} + \sum_{k=1}^j \mathbf{A}_{nk} \Phi_{n,j-k}$, $j \geq 1$, where $\mathbf{M}_{nj} = \mathbf{0}$ for $j > r$. Using these formulas one can show that the coefficient matrices, Φ_{nj} , of the MA representation satisfy (7) when denoting its elements by $b_{nj}^{(\nu, \ell)}$, and therefore the VARMA coordinate processes $Y_{ni}^{(\nu)}$, $1 \leq \nu \leq d_n$, with innovations (6) satisfy the decay condition (D).

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 [Breitung and Eickmeier, 2011], [Han and Inoue, 2015] and [Horváth and Rice, 2019], 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 $\mathbf{b} = \{b_{nj}^{(\nu)} : j \geq 0, \nu \in \{1, \dots, d_n\}, n \geq 1\}$ and $\mathbf{c} = \{c_{nj}^{(\nu)} : j \geq 0, \nu \in \{1, \dots, d_n\}, n \geq 1\}$ be two different coefficient arrays satisfying the decay assumption and put

$$\Sigma_{n0} = \Sigma_n(\mathbf{b}) = \mathbb{V}(\mathbf{Y}_n(\mathbf{b})), \quad \Sigma_{n1} = \Sigma_n(\mathbf{c}) = \mathbb{V}(\mathbf{Y}_n(\mathbf{c})).$$

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

$$(8) \quad \mathbf{b} \neq \mathbf{c} \Rightarrow \Sigma_{n0} \neq \Sigma_{n1}, \quad n \in \mathbb{N}.$$

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 Σ_{n0} to Σ_{n1} . Let $\mathcal{V}_n = \{(\mathbf{x}_n, \mathbf{y}_n) \in \mathbb{R}^{d_n} \times \mathbb{R}^{d_n} : \mathbf{x}_n^\top \Sigma_{n0} \mathbf{y}_n \neq \mathbf{x}_n^\top \Sigma_{n1} \mathbf{y}_n\}$. 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

$$(9) \quad \mathbf{Y}_{ni} = \mathbf{Y}_{ni}(\mathbf{b})\mathbf{1}(i \leq \tau) + \mathbf{Y}_{ni}(\mathbf{c})\mathbf{1}(i > \tau), \quad 1 \leq i \leq n,$$

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

$$(10) \quad \tau = \lfloor n\vartheta \rfloor,$$

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

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

which implies a change in the second moment structure of the vector time series $\mathbf{Y}_{n1}, \dots, \mathbf{Y}_{nn}$ when H_1 holds, and in estimation of the change-point to locate the change. Under the null hypothesis the covariance matrix of \mathbf{Y}_{ni} , $1 \leq i \leq n$, is given by $\mathbb{V}(\mathbf{Y}_{n1}(\mathbf{b})) = \Sigma_{n0}$, whereas it changes under the alternative hypothesis from Σ_{n0} to $\mathbb{V}(\mathbf{Y}_{n,\tau+1}(\mathbf{c})) = \Sigma_{n1}$. If $(\mathbf{v}_n, \mathbf{w}_n) \in \mathcal{V}_n$, then the change is present in the sequence of the associated quadratic forms, $\sigma_n^2[k] = \mathbf{v}_n^\top \mathbb{V}(\mathbf{Y}_{nk}) \mathbf{w}_n$, $1 \leq k$, which change from $\mathbf{v}_n^\top \Sigma_{n0} \mathbf{w}_n$ to $\mathbf{v}_n^\top \Sigma_{n1} \mathbf{w}_n$ 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

$$(11) \quad \inf_{n \geq 1} |\Delta_n| > 0, \quad \Delta_n := \mathbf{v}_n^\top \Sigma_{n0} \mathbf{w}_n - \mathbf{v}_n^\top \Sigma_{n1} \mathbf{w}_n.$$

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 $\mathbf{Y}_{ni} \mathbf{Y}_{ni}^\top$,

$$\mathbf{S}_{nk} = \sum_{i \leq k} \mathbf{Y}_{ni} \mathbf{Y}_{ni}^\top, \quad k \geq 1,$$

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

$$U_{nk} = \mathbf{v}_n^\top \mathbf{S}_{nk} \mathbf{w}_n, \quad k \geq 1, n \geq 1.$$

Consider the CUSUM-type statistic,

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

The large sample approximations for C_n obtained in [Steland and von Sachs, 2017] 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

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

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

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

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 to use Gumbel-type extreme value asymptotics, [Csörgő and Horváth, 1997], known to

converge slowly, see, e.g., [Ferber, 2018]. For a discussion of the class of proper weight functions ensuring that $B^0(t)/g(t)$ is a.s. finite we refer to [Csörgő and Horváth, 1993]. One may use the weight function $[t(1-t)]^\beta$ for some $0 < \beta < 1/2$ or any weight function g satisfying

$$(13) \quad g(t) \geq C_g [t(1-t)]^\beta, \quad 0 \leq t \leq 1, \quad \text{for some constant } C_g.$$

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

$$(14) \quad T_n(g) > q_g(1-\alpha),$$

where q_g denotes the quantile function of the law of $\sup_{0 < t < 1} |B^0(t)|/g(t)$. As studied in [Ferber, 2018], one may also standardize the unweighted CUSUM statistic by its maximizing point, i.e., substitute $g(k/n)$ by $\sqrt{\hat{\tau}_n(1-\hat{\tau}_n)}$. The associated Brownian bridge standardized by its argmax attains a density which has been explicitly calculated in [Ferber, 2018].

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, $\tilde{\mathbf{S}}_{nk} = \sum_{i \leq k} (\mathbf{Y}_{ni} - \bar{\mathbf{Y}}_n)(\mathbf{Y}_{ni} - \bar{\mathbf{Y}}_n)^\top$, where $\bar{\mathbf{Y}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_{ni}$. The associated weighted CUSUM statistics are then given by

$$\tilde{C}_n(g) = \max_{1 \leq k < n} \frac{1}{\sqrt{n}g(k/n)} \left| \mathbf{v}_n^\top \left(\tilde{\mathbf{S}}_{nk} - \frac{k}{n} \tilde{\mathbf{S}}_{nn} \right) \mathbf{w}_n \right|, \quad \tilde{T}_n(g) = \frac{\tilde{C}_n(g)}{\hat{\alpha}_n(\mathbf{b})},$$

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

To estimate the unknown change-point τ , we propose to use the estimator

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

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

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

$$\mathbf{C}_n = (C_n(\mathbf{v}_{nj}, \mathbf{w}_{nj}))_{j=1}^L, \quad \mathbf{T}_n = (T_n(\mathbf{v}_{nj}, \mathbf{w}_{nj}))_{j=1}^L.$$

Observe that this transform differs from the transform studied in [Wang and Samworth, 2018], 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 $\{\mathbf{v}_n, \mathbf{w}_n\}$

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

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

$$H_1 : \exists j \in \{1, \dots, L\} : \mathbf{v}_{nj}^\top \mathbb{V}(\mathbf{Y}_{n\tau}) \mathbf{w}_{nj} \neq \mathbf{v}_{nj}^\top \mathbb{V}(\mathbf{Y}_{n,\tau+1}) \mathbf{w}_{nj}.$$

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

$$(15) \quad Q_n = (\mathbf{T}_n - \boldsymbol{\mu}_n^*)^\top (\hat{\boldsymbol{\Sigma}}_n^B)^- (\mathbf{T}_n - \boldsymbol{\mu}_n^*) > q_{mv}(1-\alpha).$$

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, $\boldsymbol{\mu}_n^* = \left(\max_{1 \leq k < n} \mathbb{E} \max_{1 \leq k < n} |\bar{B}^0(k/n)/g(k/n)| \right)_{j=1}^L$,

$(\hat{\boldsymbol{\Sigma}}_n^T)^-$ is the Moore-Penrose generalized inverse of $\hat{\boldsymbol{\Sigma}}_n^T = \left(\hat{\beta}_n^2(j, k) \hat{\beta}_n^{-1}(k, k) \hat{\beta}_n^{-1}(j, j) \right)_{\substack{1 \leq j \leq L \\ 1 \leq k \leq L}}$ and

$q_{mv}(p)$ denotes the p -quantile of the simulated distribution of Q_n using a Monte Carlo estimate of $\mathbb{E} \max_{1 \leq k < n} |\bar{B}^0(k/n)/g(k/n)|$; the estimators $\hat{\beta}_n^2(j, k)$ of the asymptotic covariance of the j th and k th coordinate of the CUSUM transform \mathbf{C}_n are defined in the next section, 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 $\operatorname{span}\{\mathbf{v}_{n1}, \dots, \mathbf{v}_{nL}\}$ by putting $\mathbf{w}_{ni} = \mathbf{v}_{ni}$, $i \in \{1, \dots, n\}$.

2.6. Choice of the projections. The question arises how to choose the projection vectors $\mathbf{v}_n, \mathbf{w}_n$. Their choice may depend on the application. Here are some examples.

Example 1. (*Change of sets of covariances as in gene expression time series*)

Time series gene expression studies investigate the gene expression levels of a large number of genes measured at several time points, in order to identify and analyze activated genes and their relationship in a biological process, see [Bar-Joseph et al., 2012]. Going beyond the expression levels and analyzing the dependence structure of gene expression is of interest. For example, a group of genes may be uncorrelated to others or the rest of the genome, but interactions inducing correlations may start after an external stimulus. To analyze two groups, e.g., the first p and the last q variables, one may use $\mathbf{v}_n = q^{-1}(\mathbf{1}_q^\top, \mathbf{0}_{d_n-q}^\top)^\top$ and $\mathbf{w}_n = p^{-1}(\mathbf{0}_{d_n-p}^\top, \mathbf{1}_p^\top)^\top$, corresponding to $\sigma_n^2[k] = \frac{1}{pq} \sum_{j=1}^q \sum_{\ell=d_n-p+1}^{d_n} \mathbb{C}(Y_{nk}^{(j)}, Y_{nk}^{(\ell)})$, the average covariance between the first q and the last p coordinates. Further, in order to compare the first q variables with the remaining ones, one could use $\mathbf{w}_n = (\mathbf{0}_q^\top, (1/2)^r, (1/3)^r, \dots, 1/(d_n - q + 1)^r)^\top$ with $r > 1$.

Example 2. (*Spatial clustered sensors*)

Suppose that the d_n observed variables represent sensors of r clusters or groups, e.g., sensors spatially distributed over r geographic regions such as states. Such a classification is given by a partition $\cup_{i=1}^r \mathcal{J}_{ni} = \{1, \dots, d_n\}$ with pairwise disjoint sets $\emptyset \neq \mathcal{J}_{ni}$, $i = 1, \dots, r$. To analyze the within-region and between-region covariance structures, one may consider the orthogonal system given by the vectors $\mathbf{v}_{ni} = |\mathcal{J}_{ni}|^{-1}(\mathbf{1}_{\mathcal{J}_{ni}}(\nu))_{\nu=1}^{d_n}$, $i = 1, \dots, r$. Our results allow for the case of region-wise infill asymptotics where $|\mathcal{J}_{ni}|$ increases with the sample size. In our data example, the grouping is, however, determined by a sparse PCA instead of using geographic locations.

Example 3. (*Change in the autocovariance function (ACVF) of a stationary time series*)

Our high-dimensional time series model also allows to analyze the ACVF of a stationary time series. Let $X_i^{(n)} = \sum_{j=0}^{\infty} c_{nj} \epsilon_{i-j}$ be a stationary linear time series with coefficients $\{c_{nj} : j \geq 0\}$, $n \geq 1$, satisfying Assumption (D) and define

$$Y_i^{(\nu)} = X_{n,i+\nu}, \quad i \in \mathbb{N}, \nu = 1, \dots, n - d_n.$$

Then $\mathbf{Y}_{ni} = (Y_{ni}^{(1)}, \dots, Y_{ni}^{(d_n)})^\top$ is a special case of model (1), and the change-point model (9) analyzes a change of the coefficients of $X_t^{(n)}$ in terms of the ACVF $\gamma_n(h) = \mathbb{E}(X_1^{(n)} X_{1+h}^{(n)})$ up to the lag d_n respectively a change of the ACVF due to a change of the underlying coefficients. Since then the sample covariance matrix consists of the sample autocovariance estimators, the proposed CUSUM tests consider a weighted averages of them and taking unit vectors for $\mathbf{v}_n, \mathbf{w}_n$ leads to a procedure closely related to the CUSUM test studied in [Berkes et al., 2009]. Changes in autocovariances have also been studied by [Na et al., 2011] from a parametric point of view and by [Steland, 2005] and [Li and Zhao, 2013] using kernel methods.

Example 4. (*Financial portfolio analysis*)

In financial portfolio optimization one is given a stationary time series of returns \mathbf{Y}_{nt} of d_n assets and seeks a portfolio vector representing the number of shares to hold from each asset. The variance-minimizing portfolio \mathbf{w}_n^* is obtained by minimizing the portfolio risk $\mathbb{V}(\mathbf{w}_n^\top \mathbf{Y}_n)$ under the constraint $\mathbf{1}^\top \mathbf{w}_n = 1$. In order to keep transactions costs moderate, sparsity constraints can be added, see, e.g., [Brodie et al., 2009] where a ℓ_1 -penalty term is added. For bounds and confidence intervals of the risk $\mathbf{w}_n^{*\top} \Sigma_n \mathbf{w}_n^*$ of the optimal portfolio see [Steland, 2018].

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) $\mathbf{v}_n^{(1)}, \dots, \mathbf{v}_n^{(K)}$. 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 $\mathbf{v}_n^{(i)\top} \mathbf{Y}_n$, $i \in \{1, \dots, K\}$, for simplicity also called projections, and represent a lower dimensional compressed approximation of \mathbf{Y}_n . The uncertainty of its coordinates, i.e., of its position in the subspace, can be measured by the variances $\mathbf{v}_n^{(i)\top} \Sigma_n \mathbf{v}_n^{(i)}$. 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 $\mathbf{v}_n^{(i)\top} \boldsymbol{\Sigma}_n \mathbf{v}_n^{(j)}$, $1 \leq i, j \leq K$. Also observe that one may analyze the spectrum, since for eigenvectors $\mathbf{v}_n^{(i)}$ the associated eigenvalue is given by $\mathbf{v}_n^{(i)\top} \boldsymbol{\Sigma}_n \mathbf{v}_n^{(i)}$.

The question under which conditions PCA or sparse PCA is consistent has been studied by various authors. The classic Davis-Kahan theorem, see [Davis and Kahan, 1970] and [Yu et al., 2015] 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 $d/n \rightarrow y \in (0, \infty]$, 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 [Johnstone and Lu, 2009] if $d = o(n)$, [Paul, 2007] under the condition $d/n \rightarrow \gamma \in (0, 1)$ and [Jung and Marron, 2009] for n fixed and $d \rightarrow \infty$. Sparse principal components, first formally studied by [Jolliffe et al., 2003] 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 [Shen et al., 2013] for simple thresholding sparse PCA when n is held fixed and $d \rightarrow \infty$, [Birnbaum et al., 2013] for results on minimax rates when estimating the leading eigenvectors under ℓ_q -constraints on the eigenvectors and fixed eigenvalues, whereas [Cai et al., 2015] provide minimax bounds assuming at most k entries of the eigenvectors are non-vanishing and [Wang and Fan, 2017] derives asymptotic distributions allowing for diverging eigenvalues and $d/n \rightarrow \infty$.

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 [Johnson and Lindenstrauss, 1984]. 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 ℓ_0 -sparsity $O(\varepsilon^{-1} \log(1/\delta))$ of the associated projection matrix, see [Kane and Nelson, 2014].

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,

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

and their centered versions,

$$(17) \quad \mathbf{D}_{nk} = \mathbf{U}_{nk} - \mathbb{E}(\mathbf{U}_{nk}),$$

for $k, n \geq 0$, where for brevity $Y_{ni}(\mathbf{v}_n^\top \mathbf{b}_n)$ and $Y_{ni}(\mathbf{v}_n^\top \mathbf{c}_n)$ are defined by

$$Y_{ni}(\mathbf{z}^\top \mathbf{a}) = \sum_{j=0}^{\infty} \sum_{\nu=1}^{d_n} z_{n\nu} a_{nj}^{(\nu)} \epsilon_{i-j}$$

for $\mathbf{z} \in \{\mathbf{v}, \mathbf{w}\}$ and $\mathbf{a} \in \{\mathbf{b}, \mathbf{c}\}$.

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 \mathbf{D}_{nk} can be approximated by a \mathcal{F}_{nk} -martingale array with asymptotic covariance parameter $\beta_n^2(\mathbf{b}, \mathbf{c})$ defined in Lemma 1, see (48). Denote by $\alpha_n^2(\mathbf{a}) = \beta_n^2(\mathbf{a}, \mathbf{a})$, for $\mathbf{a} \in \{\mathbf{b}, \mathbf{c}\}$, the associated asymptotic variance parameter.

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

$$(18) \quad \begin{pmatrix} \mathbb{V}(\mathbf{B}_n^{(1)}) & \mathbb{C}(\mathbf{B}_n^{(1)}, \mathbf{B}_n^{(2)}) \\ \mathbb{C}(\mathbf{B}_n^{(1)}, \mathbf{B}_n^{(2)}) & \mathbb{V}(\mathbf{B}_n^{(2)}) \end{pmatrix} = \begin{pmatrix} \alpha_n^2(\mathbf{b}) & \beta_n^2(\mathbf{b}, \mathbf{c}) \\ \beta_n^2(\mathbf{b}, \mathbf{c}) & \alpha_n^2(\mathbf{c}) \end{pmatrix}, \quad n \geq 1.$$

For $n \geq 1$ define the Gaussian processes

$$(19) \quad \begin{aligned} G_n(t) &= \mathbf{B}_n^{(1)}(t) \mathbf{1}(t \leq \tau) + [\mathbf{B}_n^{(1)}(\tau) + (\mathbf{B}_n^{(2)}(t) - \mathbf{B}_n^{(2)}(\tau))] \mathbf{1}(t > \tau), \quad t \geq 0, \\ \bar{G}_n(t) &= \frac{1}{\sqrt{n}} G_n(tn), \quad t \in [0, 1]. \end{aligned}$$

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

$$\begin{aligned} G_n^0(k) &= G_n(k) - \frac{k}{n} G_n(n), \quad k \leq n, n \geq 1, \\ \bar{G}_n^0(t) &= \bar{G}_n(t) - t \bar{G}_n(1), \quad t \in [0, 1]. \end{aligned}$$

As shown in the appendix, it holds

$$(20) \quad \begin{aligned} \mathbb{C}(G_n(s), G_n(t)) &= \begin{cases} \min(s, t) \alpha_n^2(\mathbf{b}), & s, t \leq \tau \text{ or } s \leq \tau < t, \\ \min(s - \tau, t - \tau) \alpha_n^2(\mathbf{c}), & \tau \leq s, t. \end{cases} \\ \mathbb{C}(G_n^0(s), G_n^0(t)) &= \begin{cases} (\min(s, t) - \frac{st}{n}) \alpha_n^2(\mathbf{b}), & s, t \leq \tau \text{ or } s \leq \tau < t, \\ (\min(s - \tau, t - \tau) - \frac{st}{n}) \alpha_n^2(\mathbf{c}), & \tau \leq s, t. \end{cases} \end{aligned}$$

3.2. Change-point Gaussian approximations. Closely related to the CUSUM procedures are the following càdlàg processes: Define

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

and the introduce the associated bridge process

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

Observe that its expectation is $\mathbb{E}(\mathcal{D}_n^0(k/n)) = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^k \sigma_n^2[i] - \frac{k}{n} \sum_{i=1}^n \sigma_n^2[i] \right)$, and vanishes, if $\sigma_n^2[1] = \dots = \sigma_n^2[n]$. But a non-constant series $\sigma_n^2[i]$, $i \in \{1, \dots, n\}$, may lead to $\mathbb{E}(\mathcal{D}_n^0(k/n)) \neq 0$. 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 theses càdlàg processes for nice weighting functions g .

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

$$\begin{aligned} U_{nk} &= \mathbf{v}_n^\top \mathbf{S}_{nk} \mathbf{w}_n, \\ D_{nk} &= U_{nk} - \mathbb{E}(U_{nk}) = \mathbf{v}_n^\top (\mathbf{S}_{nk} - \mathbb{E}(\mathbf{S}_{nk})) \mathbf{w}_n, \end{aligned}$$

and

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

The following theorem extends the results of [Steland and von Sachs, 2017, Steland and von Sachs, 2018] 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 $\mathbf{v}_n, \mathbf{w}_n$ be weighting vectors with ℓ_1 -norms satisfying*

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

and let $\mathbf{b} = \{b_{nj}^{(\nu)}\}$ and $\mathbf{c} = \{c_{nj}^{(\nu)}\}$ 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 $\{\mathbf{B}_n(t) : t \in [0, 1]\}$ with coordinates $\mathbf{B}_n^{(i)}(t)$, $t \in [0, 1]$, $i = 1, 2$, characterized by the covariance matrix (18) associated to the parameters $\alpha_n^2(\mathbf{b}), \alpha_n^2(\mathbf{c})$, assumed to be bounded away from zero, and $\beta_n^2(\mathbf{b}, \mathbf{c})$, such that for some constant C_n the following assertions hold true almost surely:

- (i) $\|\mathbf{D}_{nt} - \mathbf{B}_n(t)\|_2 \leq C_n t^{1/2-\lambda}$, $t > 0$.
- (ii) $\max_{1 \leq k < n} \|\mathbf{D}_{nk} - \frac{k}{n} \mathbf{D}_{nn} - [\mathbf{B}_{nk} - \frac{k}{n} \mathbf{B}_n(n)]\|_2 \leq 2C_n n^{1/2-\lambda}$, $n \geq 1$.
- (iii) $\max_{1 \leq k < n} \frac{1}{\sqrt{n}} |D_{nk} - \frac{k}{n} D_{nn} - G_n^0(k)| \leq 6\sqrt{2}C_n n^{-\lambda}$, $n \geq 1$.
- (iv) $\left| \max_{1 \leq k < n} \frac{1}{\sqrt{n}} |D_{nk} - \frac{k}{n} D_{nn}| - \max_{1 \leq k < n} \frac{1}{\sqrt{n}} |G_n^0(k)| \right| \leq 6\sqrt{2}C_n n^{-\lambda}$, $n \geq 1$.
- (v) $\max_{1 \leq k < n} \frac{1}{\sqrt{n}} |U_{nk} - \frac{k}{n} U_{nn} - [m_n(k) + G_n^0(k)]| \leq 6\sqrt{2}C_n n^{-\lambda}$, $n \geq 1$.
- (vi) $\left| \max_{1 \leq k < n} \frac{1}{\sqrt{n}} |U_{nk} - \frac{k}{n} U_{nn}| - \max_{1 \leq k < n} \frac{1}{\sqrt{n}} |m_n(k) + G_n^0(k)| \right| \leq 6\sqrt{2}C_n n^{-\lambda}$, $n \geq 1$.

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

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

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

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

and

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

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

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), $\max_{1 \leq k < n} \frac{1}{\sqrt{n}} |U_{nk} - \frac{k}{n} U_{nn}| \xrightarrow{\mathbb{P}} \infty$, $n \rightarrow \infty$.*

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

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

If this function is asymptotically constant, especially if $\Delta_n \neq 0$ for all n but $\sqrt{n} \Delta_n = o(1)$ (which implies $|\alpha_n^2(\mathbf{b}) - \alpha_n^2(\mathbf{c})| = o(1)$ by (8) and Lemma 1) and $\tau/n \rightarrow \vartheta \in (0, 1)$, then the change is asymptotically not detectable, since the asymptotic law is the same as under the null hypothesis.

Now assume $\tau/n \rightarrow \vartheta$. A change ϑ located in a measurable set $A \subset (0, 1)$ with positive Lebesgue measure is detectable and changes the asymptotic law, if $H_n \rightarrow h^*$, $n \rightarrow \infty$, for some function $h^* \neq 0$ on A , since then the asymptotic law is given by $\sup_{0 < t < 1} |[h^*(t) + B^0(t)]/g(t)|$, or if $H_n \xrightarrow{\mathbb{P}} \infty$, $n \rightarrow \infty$, on $[\vartheta, 1)$, cf. Theorem 2. The case $H_n \rightarrow h^*$ corresponds to a local alternative such as $\Sigma_{n1} = \Sigma_{n0} + \Delta_n/\sqrt{n}$ for some $d_n \times d_n$ matrix Δ_n such that $\lim_{n \rightarrow \infty} \mathbf{v}_n^\top \Delta_n \mathbf{w}_n$ exists. For example, if in the spiked covariance model (3) a new local spike term of the form $n^{-1/2} \lambda_{r+1} \mathbf{u}_{n,r+1}$ appears after the change-point, then $\Delta_n = \lambda_{r+1} \mathbf{u}_{n,r+1}$ and $\Delta_n = \lambda_{r+1} \mathbf{v}_n^\top \mathbf{u}_{n,r+1} \mathbf{w}_n^\top \mathbf{u}_{n,r+1}$. 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 $\Delta_n = \mathbf{v}_n^\top (\Sigma_{n0} - \Sigma_{n1}) \mathbf{w}_n$. Clearly, $|\Delta_n|$ is maximized if $\mathbf{v}_n = \mathbf{w}_n$ is a leading eigenvector of $\Sigma_{n0} - \Sigma_{n1}$. This can be seen from the spectral decomposition $\Delta_n = \sum_{i=1}^s \phi_{ni} \delta_{ni} \delta_{ni}^\top$, where δ_{ni} are the eigenvectors and ϕ_{ni} the eigenvalues. When there is no knowledge about the change, e.g., in terms of the ϕ_{ni} and/or δ_{ni} or in terms of the model coefficients $c_{nj}^{(\nu)}$, it makes sense to select $\mathbf{v}_n, \mathbf{w}_n$ from a known basis or as leading (sparse) eigenvectors of Σ_{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 $\sup_{n \geq 1} \max_{1 \leq \nu \leq d_n} |c_{nj}^{(\nu)}| \ll (j \vee 1)^{-1-\theta}$ for some $\theta > 0$ hold. Suppose that the vector time series is centered at the sample averages $\hat{\mu}_\nu = \frac{1}{n} \sum_{i=1}^n Y_{ni}^{(\nu)}$, before applying the CUSUM procedures, leading to the statistics $\tilde{C}_n(g)$ and $\tilde{T}_n(g)$. Then assertions (i) and (ii) of Theorem 1 hold true with an additional error term $o_{\mathbb{P}}(n^{1/2})$ and (iii)-(vi) with an additional $o_{\mathbb{P}}(1)$ term. Finally, (vii) and (viii) hold in probability, if $C_n n^{-\lambda} = o(1)$.*

The above theorems assume that the projection vectors \mathbf{v}_n and \mathbf{w}_n have uniformly bounded ℓ_1 -norm. When standardizing by a homogenous estimator $\hat{\alpha}_n = \hat{\alpha}_n(\mathbf{v}_n, \mathbf{w}_n)$, i.e. satisfying

$$(24) \quad \hat{\alpha}_n(x\mathbf{v}_n, y\mathbf{w}_n) = xy\hat{\alpha}_n(\mathbf{v}_n, \mathbf{w}_n)$$

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

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

$$(25) \quad \sup_{n \geq 1} d_n^{-1/2} \|\mathbf{v}_n\|_{\ell_2}, \sup_{n \geq 1} d_n^{-1/2} \|\mathbf{w}_n\|_{\ell_2} < \infty$$

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

$$(26) \quad \sup_{n \geq 1} a_n^{-1} \|\mathbf{v}_n\|_{\ell_1}, \sup_{n \geq 1} b_n^{-1} \|\mathbf{w}_n\|_{\ell_1} < \infty.$$

Suppose that the estimator $\hat{\alpha}_n = \hat{\alpha}_n(\mathbf{v}_n, \mathbf{w}_n)$ used by $T_n(g; \mathbf{v}_n, \mathbf{w}_n)$ is ratio consistent and homogenous. Further, let $\mathbf{b} = \{b_{nj}^{(\nu)}\}$ and $\mathbf{c} = \{c_{nj}^{(\nu)}\}$ 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)

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

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

By Theorem 1, statistical properties of the CUSUM statistic $C_n(g)$ can be approximated by those of $\max_{1 \leq k < n} \frac{1}{\sqrt{ng(k/n)}} |m_n(k) + G_n^0(k)|$. 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 τ and $(\min(s, t) - st)\alpha_n^2(\mathbf{c})/\alpha_n^2(\mathbf{b})$ 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 $|\alpha_n^2(\mathbf{b}) - \alpha^2(\mathbf{c})| = o(1)$ by (8) and Lemma 1. Then the asymptotics of the change-point procedures is governed by a standard Brownian bridge. Theorems 1, 4 and 3 (under the strengthened decay condition) imply FCLTs.

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

$$\mathcal{D}_n^0 \Rightarrow \mu + \bar{G}^0, \quad n \rightarrow \infty,$$

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 process \bar{G}^0 defined on $[0, 1]$ with $\mathbb{C}(\bar{G}^0(s), \bar{G}^0(t)) = (\min(s, t) - st)\alpha^2(\mathbf{b})$ if $s, t \leq \vartheta$ or $s \leq \vartheta < t$, and $\mathbb{C}(\bar{G}^0(s), \bar{G}^0(t)) = (\min(s - \vartheta, t - \vartheta) - st)\alpha^2(\mathbf{c})$, if $\tau \leq s, t$. Further, if $\mathbf{v}_n, \mathbf{w}_n$ are weighting vectors satisfying (21), (25) or (26) and if the constructions of Theorem 1 and Theorem 4, respectively, hold with $C_n n^{-\lambda} = o(1)$, then for any weight function g which satisfies (13) we have

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

3.3. Multivariate CUSUM approximation. Let us now consider $L = L_n \in \mathbb{N}$ CUSUM statistics $\mathbf{C}_n(g) = (C_{n1}(g), \dots, C_{nL_n}(g))^\top$ where

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

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

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

$$(28) \quad \Sigma_n^B = (\Sigma_{nij}^B)_{\substack{1 \leq i \leq L_n \\ 1 \leq j \leq L_n}}$$

with blocks

$$\Sigma_{nij}^B = \begin{pmatrix} \beta_n^2(\mathbf{b}, i, \mathbf{b}, j) & \beta_n^2(\mathbf{b}, i, \mathbf{c}, j) \\ \beta_n^2(\mathbf{c}, i, \mathbf{b}, j) & \beta_n^2(\mathbf{c}, i, \mathbf{c}, j) \end{pmatrix}, \quad 1 \leq i, j \leq L_n,$$

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

$$\mathbf{G}_n(t) = \mathbf{B}_n^{(1)} \mathbf{1}(t \leq \tau) + [\mathbf{B}_n^{(1)}(\tau) + (\mathbf{B}_n^{(2)}(t) - \mathbf{B}_n^{(2)}(\tau)) \mathbf{1}(t > \tau), \quad t \geq 0,$$

$$\mathbf{G}_n^0(k) = \mathbf{G}_n(k) - \frac{k}{n} \mathbf{G}_n(n), \quad k \leq n, n \geq 1,$$

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

Theorem 6. Suppose that $\{\epsilon_{ni} : i \in \mathbb{Z}, n \geq 1\}$ satisfies Assumption (E). Let $\mathbf{v}_{nj}, \mathbf{w}_{nj}$, $j \in \{1, \dots, L_n\}$, be weighting vectors satisfying (21) uniformly in j , and let $\mathbf{b} = \{b_{nj}^{(\nu)}\}$ and $\mathbf{c} = \{c_{nj}^{(\nu)}\}$ 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 $\mathbf{B}_n = (\mathbf{B}_{nj})_{j=1}^{2L_n}$ with covariance function given by (28), such that

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

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

$$(30) \quad \max_{j \leq L_n} \mathbb{P} \left(\left| L_n^{-1/2} C_n(\mathbf{v}_{nj}, \mathbf{w}_{nj}; g) - \max_{1 \leq k < n} \frac{1}{\sqrt{ng(k/n)}} |m_{nj}(k) - \mathbf{G}_{nj}^0(k)| \right| > \delta \right) = o(1),$$

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

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

3.4. Full-sample and stopped-sample estimation of $\alpha_n^2(\mathbf{b})$ and $\beta_n^2(\mathbf{b}, \mathbf{c})$. Let us now discuss how to estimate the parameter $\alpha_n^2(\mathbf{b})$ for one pair $(\mathbf{v}_n, \mathbf{w}_n)$ of projection vectors, which is used in the change-point test statistic for standardization, and the asymptotic covariance parameters $\beta_n^2(j, k) = \beta_n^2(\mathbf{b}, \mathbf{v}_{nj}, \mathbf{w}_{nj}, \mathbf{b}, \mathbf{v}_{nk}, \mathbf{w}_{nk})$ for two pairs $(\mathbf{v}_{nj}, \mathbf{w}_{nj})$ and $(\mathbf{v}_{nk}, \mathbf{w}_{nk})$, which arise in the multivariate test for a set of projections. If there is no change, one may use the proposal of [Steland and von Sachs, 2017]. 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, $\mathbf{Y}_{n1}, \dots, \mathbf{Y}_{nk}$, 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 $u \in [\varepsilon, 1]$

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

where

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

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

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

with

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

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

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 [Steland and von Sachs, 2017].

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

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

- (i) $w_{mh} \rightarrow 1$, as $m \rightarrow \infty$, for all $h \in \mathbb{Z}$, and
- (ii) $0 \leq w_{mh} \leq W < \infty$, for some constant W , for all $m \geq 1$, $h \in \mathbb{Z}$.

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

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

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

$$\sup_{u \in [\varepsilon, 1]} |\hat{\alpha}_n^2(u) - \alpha^2(u)| \xrightarrow{\mathbb{P}} 0,$$

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

$$\sup_{u \in [\varepsilon, 1]} |\hat{\beta}_n^2(u, j, k) - \beta^2(u, j, k)| \xrightarrow{\mathbb{P}} 0,$$

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

Let us now suppose we are given a consistent estimator $\hat{\tau}_n$ of the unknown change-point; in the next section we make a concrete proposal. In order to estimate the parameter $\alpha^2(\mathbf{b})$ 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 $\alpha^2(\mathbf{b})$ by $\hat{\alpha}_{\hat{\tau}_n}^2$. The following result shows that this estimator is consistent under weak conditions.

Theorem 8. Suppose that $\hat{\tau}_n$ is an estimator of τ satisfying $\hat{\tau}_n/n \in [\varepsilon, 1]$ a.s. and $|\frac{\hat{\tau}_n}{n} - \vartheta| \xrightarrow{\mathbb{P}} 0$, as $n \rightarrow \infty$. Then

$$|\hat{\alpha}_{\hat{\tau}_n}^2 - \alpha^2(\mathbf{b})| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

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 $\hat{\tau}_n$ by

$$\hat{\tau}_n = \operatorname{argmax}_{1 \leq k \leq n} |\hat{\mathcal{U}}_n(k)|, \quad \hat{\mathcal{U}}_n(k) = \frac{1}{g(k/n)n} \left(U_{nk} - \frac{k}{n} U_{nn} \right), \quad 1 \leq k \leq n, n \geq 1.$$

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

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

$$(33) \quad \Delta = \lim_{n \rightarrow \infty} \Delta_n, \quad i = 0, 1,$$

exists. To proceed, we need further notation. Put

$$(34) \quad \mathcal{U}_n(k) = \mathbb{E}(\hat{\mathcal{U}}_n(k)) = \begin{cases} \frac{k(n-\tau)}{g(k/n)n^2} \Delta_n, & k \leq \tau, \\ \tau \frac{n-k}{g(k/n)n^2} \Delta_n, & k > \tau, \end{cases}$$

and introduce the associated rescaled functions

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

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

and

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

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

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

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 $|\hat{u}_n(t)|$, $t \in [0, 1]$, converge to the true change-point ϑ . But the maximizers $\hat{\tau}_n$ of $|\hat{\mathcal{U}}_n|$ and \hat{t}_n of $|\hat{u}_n|$ are related by

$$(39) \quad \hat{\tau}_n = \operatorname{argmax}_{1 \leq k \leq n} \hat{\mathcal{U}}_n(k) = \operatorname{argmax}_{1 \leq k \leq n} \hat{u}_n(k/n) = n \operatorname{argmax}_{t \in \{1/n, \dots, 1\}} \hat{u}_n(t) = n \hat{t}_n$$

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

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*

$$(40) \quad \max_{1 \leq k < n} |\widehat{\mathcal{U}}_n(k) - \mathcal{U}_n(k)| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty,$$

$$(41) \quad \sup_{t \in [0, 1]} |\widehat{u}_n(t) - u(t)| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

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*

$$\frac{\widehat{\tau}_n}{n} \xrightarrow{\mathbb{P}} \vartheta, \quad n \rightarrow \infty.$$

5. SIMULATIONS

To investigate the statistical performance of the change-point tests a change from a family of $\text{AR}(\rho_\nu)$ series to a family of (shifted) $\text{MA}(r)$ 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

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

with $\rho_\nu = 0.5\nu/d$, for $\nu \in \{1, \dots, d\}$ and $n \geq 1$, i.i.d. standard normal ϵ_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, α_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 $\mathbf{v}_n = \mathbf{w}_n = (1/d, \dots, 1/d)^\top$. Random projections were generated by drawing from a Dirichlet distribution, such that the projections have unit ℓ_1 norm and expectation $d^{-1} \mathbf{1}$, in order to study the effect of random perturbations around the fixed projections.

Table 2 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.

The accuracy and power of the global test related to the CUSUM transform was examined for a change to a MA model after half of the sample for the sample size $n = 500$. The design of this study is data-driven as the principal directions calculated for the ozone data set were used in addition to random projections. The global test based on the weighted CUSUM transform using the weight function $g(t) = [t(1-t)]^\beta$ with $\beta = 0.3$ was fed with the first r projections for various

values of r . The results are provided in Table 1. Each entry is based on 1,000 runs. According to these figures, the proposed global test is accurate in terms of the significance level and quite powerful.

r	level	power
2	0.045	0.9
3	0.05	0.85
4	0.042	0.9
7	0.037	0.884

Table 1. Simulated level and power of the global test based on the CUSUM transform.

6. DATA EXAMPLE

To illustrate the proposed methods, we analyze $n = 1826$ daily observations of 8 hour 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 [Schweinberger et al., 2017], namely residuals obtained after fitting cubic splines to the log-transformed data, in order to correct level and seasonal ups and downs.

The data of the first year was used to calculate a sparse PCA. We use the method of [Erichson et al., 2018] to get sparse directions \mathbf{v}_i instead of [Cai et al., 2015], since, according to the latter authors, their estimators leading to minimax rates are computationally infeasible. The sparse PCA was conducted as follows: Denote the 365×444 data matrix by \mathbf{X} . [Erichson et al., 2018] propose to calculate an orthonormal matrix \mathbf{A} and a sparse matrix $\mathbf{B} = (\mathbf{v}_1, \dots, \mathbf{v}_d)$ solving

$$\min_{\mathbf{A}, \mathbf{B}} (1/2) \|\mathbf{X} - \mathbf{XBA}^\top\|_F^2 + \psi(\mathbf{B}), \quad \mathbf{A}^\top \mathbf{A} = \mathbf{I},$$

where we used an elastic net regularization $\psi(\mathbf{B}) = \lambda_1 \|\mathbf{B}\|_{\ell_1} + \lambda_2 \|\mathbf{B}\|_{\ell_2}$ with parameters $\lambda_1 = 0.025$, $\lambda_2 = 0.1$. This analysis shows that the supports $\mathcal{S}_i = \{j : \mathbf{v}_{ij} \neq 0\}$ of the leading six projections, where $\mathbf{v}_i = (v_{i1}, \dots, v_{id})^\top$ for $i \in \{1, \dots, 6\}$, correspond to a spatial segmentation which eases interpretation. Figure 1 shows the geographic locations of these supports.

The data of the years 2011 to 2014, providing the test sample $\mathbf{Y}_{n1}, \dots, \mathbf{Y}_{nn}$ 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 $\mathbb{C}(\mathbf{v}_k^\top \mathbf{Y}_{ni}, \mathbf{v}_\ell^\top \mathbf{Y}_{ni})$, $i \in \{1, \dots, n\}$, for $1 \leq k \leq \ell \leq 7$. The asymptotic variance parameter was estimated using both the full in-sample and stopped-sample approach. The application of the unweighted CUSUM approach revealed no significances at the usual levels.

Table 2. 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								
Method	ϑ	10	100	200	ϑ	10	100	200
	Unweighted CUSUM				Weighted CUSUM			
	ϑ	10	100	200	ϑ	10	100	200
L=500	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
	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
stopped-sample	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
	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
in-sample	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
	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			
	ϑ	10	100	200	ϑ	10	100	200
	ϑ	10	100	200	ϑ	10	100	200
$L = 500$	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
	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
stopped-sample	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
	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
in-sample	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
	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

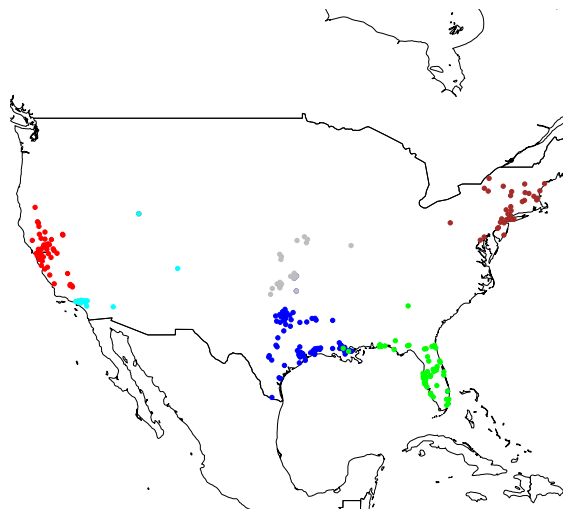


Figure 1. Sparse principal component analysis of ozone residuals from 444 monitors across the U.S.. Depicted are the locations of the supports $\mathcal{S}_1, \dots, \mathcal{S}_6$ of the leading principal directions $\mathbf{v}_1, \dots, \mathbf{v}_6$.

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 [Steland and von Sachs, 2017].

7.1. Preliminaries. For an arbitray array of coefficients $\mathbf{a} = \{a_{nj}^{(\nu)} : j \geq 0, 1 \leq \nu \leq d_n, n \geq 1\}$ and vectors $\mathbf{v}_n = (v_{n1}, \dots, v_{nd_n})^\top$ and $\mathbf{w}_n = (w_{n1}, \dots, w_{nd_n})^\top$ with finite ℓ_1 -norm, i.e., $\|\mathbf{v}_n\|_{\ell_1}, \|\mathbf{w}_n\|_{\ell_1} < \infty$, define

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

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

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

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

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

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

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

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

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

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

for $k, n \geq 1$, which takes the lag n' forward difference at m' . Notice that for $m' = 0$

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

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

$$\begin{aligned} (43) \quad & \mathbb{C}(\delta M_{m'+n'}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n), \delta M_{m'+n'}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)) \\ &= \tilde{f}_{0,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{0,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \sum_{j=1}^{n'} (\gamma_{n,m'+j} + \sigma_{n,m'+j}^4) \\ & \quad + \sum_{j=1}^{n'} \sum_{\ell=1}^{\infty} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \sigma_{n,m'+j}^2 \sigma_{n,m'+j-\ell}^2, \end{aligned}$$

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, $\beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$, using different pairs $(\mathbf{v}_n, \mathbf{w}_n)$ and $(\tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$ of weighting vectors, abbreviated as $\beta_n^2(\mathbf{b}, \mathbf{c}) = \beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \mathbf{v}_n, \mathbf{w}_n)$. Especially, it follows from these results that the asymptotic variance of a single CUSUM detector under the no-change null hypothesis, $\alpha^2(\mathbf{a}) = \beta_n^2(\mathbf{a}, \mathbf{a})$, satisfies

$$\alpha_n^2(\mathbf{a}) \approx \frac{1}{n} \mathbb{V}(D_{nn}).$$

The following general results hold under a mild condition on the error terms and especially show that (73) can be approximated by $n' \beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$ at the rate $(n')^{1-\theta}$, uniformly in n and m' , cf. [Steland and von Sachs, 2017, (3.18)] and [Kouritzin, 1995]. The proof extends these latter results and improves the bounds, but it is technical and thus deferred to the appendix. The improved bounds show that the ℓ_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 ϵ_{ni} , $i \in \mathbb{Z}$, be independent with variances σ_{ni}^2 and third moments γ_{ni} satisfying*

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

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

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 $K_n = \|\mathbf{v}_n\|_{\ell_1} \|\mathbf{w}_n\|_{\ell_1} \|\tilde{\mathbf{v}}_n\|_{\ell_1} \|\tilde{\mathbf{w}}_n\|_{\ell_1}$,

$$(46) \quad \left| \mathbb{C}(M_{n'}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n), M_{n'}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)) - (n') \beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right| \stackrel{n, n'}{\ll} K_n (n')^{1-\theta},$$

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

$$(47) \quad \left| \mathbb{C}(\delta M_{m'+n'}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n), \delta M_{m'+n'}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)) - (n') \beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right| \stackrel{n, n', m'}{\ll} K_n (n')^{1-\theta},$$

if

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

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

$$(49) \quad \mathbb{E}(S_{n', m'}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) - \delta M_{m'+n'}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n))^2 \stackrel{n, m', n'}{\ll} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{1-\theta}.$$

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

$$(50) \quad \mathbb{E}(\mathbf{D}_{nk}^{(1)} - \delta M_k^{(n)}(\mathbf{b}))^2 \stackrel{n, k}{\ll} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 k^{1-\theta},$$

$$(51) \quad \mathbb{E}(\mathbf{D}_{nk}^{(2)} - \delta M_k^{(n)}(\mathbf{c}))^2 \stackrel{n, k}{\ll} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 k^{1-\theta},$$

such that

$$(52) \quad \mathbb{E}\|\mathbf{D}_{nk} - \delta \mathbf{M}_k^{(n)}\|_2^2 \stackrel{n, k}{\ll} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 n^{1-\theta}.$$

(50), (51) and (52) also hold (with obvious modifications), if $\mathbf{D}_{nk}^{(1)} = S_{k,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n)$ and $\mathbf{D}_{nk}^{(2)} = S_{k,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$ for two pairs of weighting vectors, where the bound in (52) then is given by $\max\{\|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2, \|\tilde{\mathbf{v}}_n\|_{\ell_1}^2 \|\tilde{\mathbf{w}}_n\|_{\ell_1}^2\} n^{1-\theta}$.

Proof. See appendix. \square

The next lemma studies the conditional covariances of the approximating martingales. It generalizes [Steland and von Sachs, 2018, 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 $\beta_n^2(\mathbf{b}, \mathbf{c})$ is as defined there. Then it holds for $m', n' \geq 0$ and $n \geq 1$ with $K_n = \|\mathbf{v}_n\|_{\ell_1} \|\mathbf{w}_n\|_{\ell_1} \|\tilde{\mathbf{v}}_n\|_{\ell_1} \|\tilde{\mathbf{w}}_n\|_{\ell_1}$*

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

and

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

Proof. See appendix. \square

7.3. Proofs of Subsection 3.2. After the above preparations, we are now in a position to show Theorem 1.

Proof of Theorem 1. Put

$$(53) \quad \boldsymbol{\xi}_i^{(n)} = \boldsymbol{\xi}_i^{(n)}(\mathbf{v}_n^\top \mathbf{b}_n, \mathbf{v}_n^\top \mathbf{c}_n) = \begin{pmatrix} Y_{ni}(\mathbf{v}_n^\top \mathbf{b}_n) Y_{ni}(\mathbf{w}_n^\top \mathbf{b}_n) - \mathbb{E}[Y_{ni}(\mathbf{v}_n^\top \mathbf{b}_n) Y_{ni}(\mathbf{w}_n^\top \mathbf{b}_n)] \\ Y_{ni}(\mathbf{v}_n^\top \mathbf{c}_n) Y_{ni}(\mathbf{w}_n^\top \mathbf{c}_n) - \mathbb{E}[Y_{ni}(\mathbf{v}_n^\top \mathbf{c}_n) Y_{ni}(\mathbf{w}_n^\top \mathbf{c}_n)] \end{pmatrix},$$

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

$$\mathbf{S}_{n',m'}^{(n)} = (S_{n',m'}^{(n)}(1), S_{n',m'}^{(n)}(2))^\top = \sum_{k=m'+1}^{m'+n'} \boldsymbol{\xi}_k^{(n)}, \quad m', n' \geq 0.$$

Introduce the conditional covariance operators

$$\mathbf{C}_{n',m'}^{(n)}(\mathbf{u}) = \mathbb{E}[\mathbf{u}^\top S_{n',m'}^{(n)} S_{n',m'}^{(n)} \mid \mathcal{F}_{n,m'}], \quad \mathbf{u} \in \mathbb{R}^2, \quad n, n', m' \geq 1,$$

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

$$\mathbf{T}^{(n)}(\mathbf{u}) = \mathbb{E}[\mathbf{u}^\top \mathbf{B}_n \mathbf{B}_n], \quad \mathbf{u} \in \mathbb{R}^2, \quad n \geq 1.$$

We shall verify [Philipp, 1986, Th. 1], namely the validity of the following conditions: For $m' \geq 0$, $n' \geq 1$,

- (I) $\sup_{j \geq 1} \mathbb{E} \|\boldsymbol{\xi}_j^{(n)}\|_2^{2+\delta} < \infty$ for some $\delta > 0$.
- (II) For some $\varepsilon > 0$ it holds

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

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

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

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

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

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

and the second component is estimated analogously. Following the arguments in [Kouritzin, 1995, p. 343], for $\delta' \in (0, 2)$ and $\chi = \delta'/2$, one can show that for $\mathbf{a} \in \{\mathbf{b}, \mathbf{c}\}$ and $\mathbf{u}_n \in \{\mathbf{v}_n^*, \mathbf{w}_n^*\}$

$$\begin{aligned} \mathbb{E}|Y_{ni}(\mathbf{u}'_n \mathbf{a}_n)|^{4+\delta'} &\leq \sup_{n,k \geq 0} \mathbb{E}|\epsilon_{nk}| \sum_{\ell=0}^{\infty} |a_{n\ell}^{(u)}|^{2(2+\chi)} \\ &\quad + \sup_{n,k \geq 0} \mathbb{E}(\epsilon_{nk}^2) \left\{ \sup_{n',k' \geq 0} \mathbb{E}(\epsilon_{n'k'}^2) \right\}^{1+\chi} \sum_{\ell=0}^{\infty} |a_{n\ell}^{(u)}|^2 \left\{ \sum_{\ell=0}^{\infty} |a_{n\ell}^{(u)}|^2 \right\}^{1+\chi}, \end{aligned}$$

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 ℓ_1 -bounded \mathbf{u}_n and $n \geq 1$, 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

$$(54) \quad \max_{j=1,2} \sup_{i \geq 1} \mathbb{E}|\boldsymbol{\xi}_{ij}^{(n)}|^{2+\delta} = O(\|\mathbf{v}_n\|_{\ell_1}^{2+\delta} \|\mathbf{w}_n\|_{\ell_1}^{2+\delta}).$$

Now Jensen's inequality yields

$$\mathbb{E}\|\boldsymbol{\xi}_i^{(n)}\|_2^{2+\delta} = 2^{1+\delta/2} \mathbb{E} \left(\frac{1}{2} \sum_{j=1,2} [\boldsymbol{\xi}_{ij}^{(n)}]^2 \right)^{1+\delta/2} \leq 2^{\delta/2} \sum_{j=1,2} \mathbb{E}|\boldsymbol{\xi}_{ij}^{(n)}|^{2+\delta} < \infty,$$

verifying (I). To show (II) recall that the martingale approximation for $\mathbf{S}_{n',m'}^{(n)} = (S_{n',m'}^{(n)}(\mathbf{b}), S_{n',m'}^{(n)}(\mathbf{c}))^\top$ is given by $\delta \mathbf{M}_{n',m'}^{(n)} = (\delta M_{m'+n'}^{(n)}(\mathbf{b}), \delta M_{m'+n'}^{(n)}(\mathbf{c}))^\top$, see Lemma 2. Using

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

it follows that

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

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

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

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

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

Therefore, (III) follows, if

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

a.s., which is shown in Lemma 3, since then assumption (21) ensures the estimate $n, n', m' \ll (n')^{-\theta'/2}$. Hence, from [Philipp, 1986], we may conclude that there exists a constant C_n and a universal constant $\lambda > 0$, such that

$$(55) \quad \|D_{nt} - B_n(t)\|_2 \leq C_n t^{1/2-\lambda} \quad t > 0,$$

a.s., which implies

$$(56) \quad |D_{nt}^{(i)} - B_n^{(i)}(t)| \leq \sqrt{2} C_n t^{1/2-\lambda}, \quad t > 0,$$

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

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

we have the following crucial representation in terms of D_{nt} ,

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

for all t . Since

$$\begin{aligned} D_{nt} - \{B_n^{(1)}(t) \mathbf{1}(t \leq \tau) + [B_n^{(1)}(\tau) + B_n^{(2)}(t) - B_n^{(2)}(\tau)]\} \\ = \left(D_{nt}^{(1)} - B_n^{(1)}(t) \right) \mathbf{1}(t \leq \tau) + \left(D_{n\tau}^{(1)} - B_n^{(1)}(\tau) + D_{nt}^{(2)} - B_n^{(2)}(t) - D_{n\tau}^{(2)} + B_n^{(2)}(\tau) \right) \mathbf{1}(t > \tau), \end{aligned}$$

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

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

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

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

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

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

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

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

as $n \rightarrow \infty$, 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 γ_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.

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

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

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

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 appendix and the Hájek-Rényi inequality for martingale differences yield for any $\delta > 0$ the tail bound

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

The first term tends to 0, as $\varepsilon \rightarrow 0$, uniformly in n , since $\beta < 1/2$. Let B^0 be a Brownian bridge and note that $\{\alpha_n^{-1}(\mathbf{b})\bar{G}_n^0(t) : 0 \leq t \leq \vartheta\} \stackrel{d}{=} \{B^0(t) : 0 \leq t \leq \vartheta\}$. Using the estimates $\sqrt{t}/t^\beta \leq (\varepsilon/\gamma_n)^{1/2-\beta}$ and $\log_2(1/t) \leq \log_2(n)$ on $t \in \mathcal{G}_n = \{1/n, \dots, \lfloor \varepsilon n/\gamma_n \rfloor/n\}$ the law of the iterated logarithm for the Brownian bridge, [Shorack and Wellner, 1986, p.72], entails for $\delta > 0$ and $\varepsilon/\gamma_n \leq \vartheta$ (thus for large n)

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

by our choice of γ_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. Combining the above estimates shows (22). \square

Proof of Theorem 2. See appendix. \square

Proof of Theorem 3. See appendix. \square

Proof of Theorem 4. Since $\tilde{\mathbf{v}}_n = a_n^{-1}\mathbf{v}_n$ and $\tilde{\mathbf{w}}_n = b_n^{-1}\mathbf{w}_n$ satisfy property (21) and

$$D_{nk}(g; a_n^{-1}\mathbf{v}_n, b_n^{-1}\mathbf{w}_n) = a_n^{-1}b_n^{-1}D_{nk}(g; \mathbf{v}_n, \mathbf{w}_n),$$

we may conclude that $T_n(g; \mathbf{v}_n, \mathbf{w}_n) = T_n(g; \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$. Consequently, all approximations for T_n carry over. In particular, we obtain under the conditions of Theorem 1, cf. (23),

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

Note that $\bar{B}_n = \alpha_n^{-1}(\mathbf{b})\bar{G}_n^0(t)$ is a standard Brownian on $[0, \vartheta]$, cf. (20), whereas the scale factor changes from 1 to $\alpha_n(\mathbf{c})/\alpha_n(\mathbf{b})$ on $(\vartheta, 1]$. This shows (27) for ℓ_1 -bounded projections. The proof for uniformly ℓ_2 -bounded projections uses the scaling $a_n = b_n = d_n$ and the fact that by Jensen's inequality gives $\|\tilde{\mathbf{v}}_n\|_{\ell_1} \leq \left(\frac{1}{d_n} \sum_{\nu=1}^{\infty} w_{n\nu}^2\right)$, where the sum is finite by assumption and the factor cancels by standardization, see also [Steland and von Sachs, 2018]. \square

Proof of Theorem 5. The conditions on g ensure that $\sup_{0 < t < 1} |B^0(t)|/g(t)$ is well defined, see [Csörgő and Horváth, 1993]. Further, $\{G_n^0(t)/\alpha^2(\mathbf{b}) : 0 \leq t \leq \tau\} \stackrel{d}{=} \{B^0(t) : 0 \leq t \leq \tau\}$ and $\{G_n^0(t)/\alpha^2(\mathbf{c}) : \tau < t \leq 1\} \stackrel{d}{=} \{B^0(t) : \tau < t \leq 1\}$ for each n . Therefore, combining these facts, 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., and the continuous mapping theorem the result follows from (23). \square

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

$$\mathbf{D}_{nk} = \begin{pmatrix} D_{nk}(\mathbf{v}_{n1}, \mathbf{w}_{n1}) \\ \vdots \\ D_{nk}(\mathbf{v}_{nL}, \mathbf{w}_{nL}) \end{pmatrix} = \sum_{i \leq k} \boldsymbol{\xi}_i^{(n)}, \quad k \geq 1, \quad \boldsymbol{\xi}_i^{(n)} = \left(\xi_{ni}^{(n)}(j) \right)_{j=1}^L.$$

Also put $\mathbf{S}_{n',m'}^{(n)} = \sum_{k=m'+1}^{m'+n'} \boldsymbol{\xi}_k^{(n)}$, $n', m' \geq 0, n \geq 1$. For sparseness of notation, we use the same symbols \mathbf{D}_{nk} , $\mathbf{S}_{n',m'}^{(n)}$ 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 $L_n \rightarrow \infty$, 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 $\|\cdot\|_\infty \leq \|\cdot\|_2$ and $\|\cdot\|_2 \leq \sqrt{L_n} \|\cdot\|_\infty$ - and Jensen's inequality yield, in view of (54),

$$\sup_{i \geq 1} \mathbb{E} \|\mathbf{L}_n^{-1/2} \boldsymbol{\xi}_i^{(n)}\|_2^{2+\delta} = \sup_{i \geq 1} \mathbb{E} \left[\frac{1}{L_n} \sum_{j=1}^{L_n} \|\boldsymbol{\xi}_i^{(n)}\|_2^2 \right]^{(2+\delta)/2} \leq \sup_{i \geq 1} \frac{1}{L_n} \sum_{j=1}^{L_n} \mathbb{E} \|\boldsymbol{\xi}_i^{(n)}\|_2^{2+\delta} < \infty,$$

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

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

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

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

and the covariance operator $\mathbf{T}^{(n)}(\mathbf{u}) = \mathbb{E}(\mathbf{u}^\top \mathbf{B}_n \mathbf{B}_n)$, $\mathbf{u} \in \mathbb{R}^{2L_n}$. 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 [Steland and von Sachs, 2018]. Denote the ν th coordinate of $\mathbf{S}_{n',m'}^{(n)}$ corresponding to the weighting vectors $\mathbf{v}_n(\nu)$ and $\mathbf{w}_n(\nu)$ by $\mathbf{S}_{n',m'}^{(n)}(\nu)$ and let

$$C_{n',m'}^{(n)}(\nu, \mu) = \mathbb{E}((L_n^{-1/2} \mathbf{S}_{n',m'}^{(n)}(\nu))(L_n^{-1/2} \mathbf{S}_{n',m'}^{(n)}(\mu) \mid \mathcal{F}_{n,m'})).$$

By Lemma 3

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

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

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

which establishes condition (III). Hence, from [Philipp, 1986], 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 \mathbf{D}_{nt} and a Brownian motion as described in the theorem

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

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

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

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 $\{n_0, \dots, n_1\}$ for

$n_i = \lfloor nt_i \rfloor$, $i = 1, 2$. For a weight function g satisfying (13) and CUSUMs taking the maximum over $\{1, \dots, n-1\}$ the required LIL tail bound and the martingale approximation used to apply the Hájek-Rényi inequality do not depend on $1 \leq j \leq L_n$ or L_n , such that

$$\max_{j \leq L_n} \mathbb{P} \left(\left| L_n^{-1/2} C_n^g(\mathbf{v}_{nj}, \mathbf{w}_{nj}) - \max_{1 \leq k \leq n} \frac{1}{\sqrt{ng(k/n)}} |m_{nj}(k) - G_{nj}^0(k)| \right| > \delta \right) = o(1),$$

for any $\delta > 0$. \square

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 $\hat{\Gamma}_n(h; d)$ the estimator $\hat{\Gamma}_n(h)$ regarding the dimension d as a formal parameter such that $\hat{\Gamma}_n(h) = \hat{\Gamma}_n(h; d)|_{d=d_n}$. In the same vain we proceed for $\hat{\beta}_n^2$ and all other statistics arising below and write $\hat{\beta}_n^2(d)$ etc. The assertion will then follow by showing that the consistency is uniform in the dimension d . By assumption $z_{ni}^{(j)} = \mathbf{v}_{nj}^\top \mathbf{Y}_{ni} \mathbf{w}_{nj}^\top \mathbf{Y}_{ni} - \mathbb{E}(\mathbf{v}_{nj}^\top \mathbf{Y}_{ni} \mathbf{w}_{nj}^\top \mathbf{Y}_{ni})$ satisfies $z_{ni}^{(j)} = \mathbf{v}_{nj}^\top \mathbf{Y}_{ni}(\mathbf{b}) \mathbf{w}_{nj}^\top \mathbf{Y}_{ni}(\mathbf{b}) =: z_{ni}^{(j)}(\mathbf{b})$ for $i \leq \tau$ and $z_{ni}^{(j)} = \mathbf{v}_{nj}^\top \mathbf{Y}_{ni}(\mathbf{c}) \mathbf{w}_{nj}^\top \mathbf{Y}_{ni}(\mathbf{c}) =: z_{ni}^{(j)}(\mathbf{c})$ if $i > \tau$. Put $\xi_{ni}^{(j)} = z_{ni}^{(j)} - \mathbb{E}(z_{ni}^{(j)})$ and again let $\xi_{ni}^{(j)}(\mathbf{b}) = \xi_{ni}^{(j)}$, if $i \leq \tau$, and $\xi_{ni}^{(j)}(\mathbf{c}) = \xi_{ni}^{(j)}$, if $\tau < i \leq n$. By Lemma 1 and Lemma 2, $\beta_n^2(j, k) = n^{-1} \mathbb{C}(\mathbf{v}_{nj}^\top \mathbf{S}_{nn} \mathbf{w}_{nj}, \mathbf{v}_{nk}^\top \mathbf{S}_{nn} \mathbf{w}_{nk}) + R_n$ with $\mathbb{E}(R_n^2) = O(n^{-\theta})$. Combining this with (59), we obtain $\beta_n^2(j, k) = \sum_{h \in \mathbb{Z}} \mathbb{E} \left(\xi_{n0}^{(j)} \xi_{n,|h|}^{(k)} \right) + R_n + o(1)$. Without loss of generality we fix $(j, k) = (1, 2)$ and show that $\sum_{h \in \mathbb{Z}} \hat{\Gamma}_n(h, 1, 2) - \sum_{h \in \mathbb{Z}} \mathbb{E}(\xi_{n0}^{(1)} \xi_{n,|h|}^{(2)}) = o(1)$, as $n \rightarrow \infty$, where

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

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

Observe that for $h \geq 0$

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

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

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

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

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

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

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

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

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

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

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

as $n \rightarrow \infty$, where

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

for $u \in [\varepsilon, 1]$. As in [Steland and von Sachs, 2017, Th. 4.4] one can show that

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

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

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

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

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

yielding the representation

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

as well as

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

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

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

The arguments used in the proof of [Steland and von Sachs, 2017, 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 $\{\xi_{ni}^{(j)} : \lfloor n\vartheta \rfloor + 1 \leq i \leq n - h\}$,

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

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

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

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

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

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

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

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 $\mathbb{E} \left(n |\widehat{\Gamma}_n(u; h, d) - \widetilde{\Gamma}_n(u; h, d)| \right) = O(1)$, uniformly over $|h| \leq m_n$, $d \in \mathbb{N}$ and $u \in [\varepsilon, 1]$. For example, for $0 \leq h \leq m_n$

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

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

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

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

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

and combine (58), (60) and (65), see the appendix for details. \square

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 (84) we have for $k \leq n/2$

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

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

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

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

$$\mathbb{P} \left(\max_{1 \leq k < n} |\widetilde{D}_{nk}| > \delta n \right) = o(1)$$

and in turn (40). Next consider

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

as $n \rightarrow \infty$, by (40). Clearly, $u_n(t) \rightarrow u(t)$ 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. \square

Proof of Theorem 10. Since $\vartheta \in (0, 1)$ is an isolated maximum of u and \hat{u}_n converges uniformly to u , the consistency follows from well known results, see, e.g., [van der Vaart, 1998], by virtue of Theorem 9 and (39). \square

APPENDIX A. ADDITIONAL RESULTS AND PROOFS

This section provides technical details of several proofs and additional auxiliary results. Especially, the proofs of the asymptotic results are based on martingale approximations which require several additional results and technical preparations. These results extend and complement the results obtained in [Steland and von Sachs, 2017].

A.1. Proofs for Subsection 7.1 (Preliminaries). Consider the multivariate linear time series of dimension $q = q_n \rightarrow \infty$,

$$\mathbf{Z}_{ni} = \sum_{j=0}^{\infty} \mathbf{B}_{nj} \mathbf{P} \mathbf{V}^{1/2} \boldsymbol{\epsilon}_{n,i-j}, \quad i \geq 1, n \in \mathbb{N},$$

with $\boldsymbol{\epsilon}_{ni} = (\epsilon_{n,i-r_1}, \dots, \epsilon_{n,i-r_{d_n}})^\top$, $i \geq 1, n \in \mathbb{N}$, $0 = r_1 < \dots < r_{d_n}$, cf. (5), which can be written as

$$\mathbf{Z}_{ni} = (Z_{ni}^{(1)}, \dots, Z_{ni}^{(d_n)})^\top, \quad Z_{ni}^{(\nu)} = \sum_{k=0}^{\infty} \left(\sum_{\ell=1}^q \mathbf{1}(k \geq r_\ell) b_{n,k-r_\ell}^{(\nu,\ell)} \right) \epsilon_{n,i-k}, \quad \nu = 1, \dots, d_n.$$

Let $\mathbf{P} \mathbf{V}^{1/2} = \sum_{i=1}^q \pi_{ni} \mathbf{l}_{ni} \mathbf{r}_{ni}^\top$ be the SVD of $\mathbf{P} \mathbf{V}^{1/2}$ with singular values π_{ni} , left singular vectors $\mathbf{l}_{ni} \in \mathbb{R}^{q_n}$ and right singular vectors $\mathbf{r}_{ni} = (r_{ni1}, \dots, r_{nid_n})^\top \in \mathbb{R}^{d_n}$, which satisfy $\|\mathbf{l}_{ni}\|_{\ell_2} = \|\mathbf{r}_{ni}\|_{\ell_2} = 1$, $i = 1, \dots, q_n$. Then $\mathbf{B}_{nj} \mathbf{P} \mathbf{V}^{1/2} = \sum_{i=1}^{q_n} \pi_{ni} \mathbf{B}_{nj} \mathbf{l}_{ni} \mathbf{r}_{ni}^\top$, and the element at position (ν, ℓ) of the latter matrix is given by $\sum_{i=1}^q \pi_{ni} \mathbf{b}_{nj,\nu}^\top \mathbf{l}_{ni} r_{ni\ell}$. Plugging the latter formula into equation (4) leads us to

$$Z_{ni}^{(\nu)} = \sum_{k=0}^{\infty} \left(\sum_{\ell=1}^{q_n} \mathbf{1}(k \geq r_\ell) \sum_{i'=1}^{q_n} \pi_{ni'} \mathbf{b}_{n,k-r_\ell,\nu}^\top \mathbf{l}_{ni'} r_{ni'\ell} \right) \epsilon_{n,i-k},$$

i.e. the coefficients $c_{nk}^{(Z,\nu)}$ of the series (4) take now the form $c_{nk}^{(Z,\nu)} = \sum_{\ell=1}^{q_n} \mathbf{1}(k \geq r_\ell) \sum_{i=1}^{q_n} \pi_{ni} \mathbf{b}_{n,k-r_\ell,\nu}^\top \mathbf{l}_{ni} r_{ni\ell}$.

Lemma 4. *Suppose that*

- (i) $\|\mathbf{l}_{nk}\|_{\ell_1}, \|\mathbf{r}_{nk}\|_{\ell_1} \stackrel{n,k}{\ll} 1$ and $\sum_{i=1}^{\infty} |\pi_{ni}| = O(1)$,
- (ii) $\sup_{n \geq 1} \sup_{1 \leq \nu, \mu} |b_{nj}^{(\nu,\mu)}| \stackrel{j,\ell}{\ll} (j + 2r_\ell)^{-3/2-\theta}$,
- (iii) $\sum_{\ell=1}^{\infty} r_\ell^{-3/4-\theta/2} = O(1)$.

Then $\sup_{n \geq 1} \max_{1 \leq \nu \leq d_n} c_{nk}^{(Z,\nu)} \stackrel{k}{\ll} k^{-3/4-\theta/2}$, i.e. (D) holds.

Proof of Lemma 4. We can assume that the constants in (i) are equal to 1. By (i) $|r_{nk\ell}| \leq 1$ for all n, k, ℓ . Using the inequality $|\mathbf{x}^\top \mathbf{y}| \leq \|\mathbf{x}\|_{\ell_1} \sup_j |y_j|$ we obtain for all $\nu \geq 1$

$$|\mathbf{b}_{k-r_\ell,\nu}^{(\nu)}{}^\top \mathbf{l}_{ni} r_{ni\ell}| \leq \|\mathbf{l}_{ni}\|_{\ell_1} |r_{ni\ell}| \sup_{1 \leq \nu, \mu} |b_{k-r_\ell}^{(\nu,\mu)}| \stackrel{k,\ell}{\ll} (k + r_\ell)^{-3/2-\theta/2}.$$

Combining

(66)

$$(k + r_\ell)^{-3/2-\theta} = r_\ell^{-3/4-\theta/2} k^{-3/4-\theta/2} \left(\frac{k}{k + r_\ell} \right)^{3/4+\theta/2} \left(\frac{r_\ell}{k + r_\ell} \right)^{3/4+\theta/2} \leq r_\ell^{-3/4-\theta/2} k^{-3/4-\theta/2}$$

with $\sum_{i=1}^{\infty} |\pi_{ni}| = O(1)$ now yields

$$\sum_{i=1}^{q_n} |\pi_{ni} \mathbf{b}_{n,k-r_\ell,\nu}^\top \mathbf{l}_{ni} r_{ni\ell}| \stackrel{k,\ell}{\ll} r_\ell^{-3/4-\theta/2} (k+r_\ell)^{-3/4-\theta/2} \sum_{i=1}^{\infty} |\pi_{ni}|.$$

Using (iii) we may conclude that the coefficients $c_{nk}^{(Z,\nu)}$ satisfy

$$\sup_{n \geq 1} \max_{\nu \geq 1} |c_{nk}^{(Z,\nu)}| \leq \sum_{\ell=1}^{\infty} \left| \sum_{i=1}^{q_n} \pi_{ni} \mathbf{b}_{n,k-r_\ell,\nu}^\top \mathbf{l}_{ni} r_{ni\ell} \right| \stackrel{k}{\ll} k^{-3/4-\theta/2} \sum_{\ell=1}^{\infty} r_\ell^{-3/4-\theta/2},$$

which completes the proof. \square

Whereas the conditions of Lemma 4 rule out eigenvectors such as $(1/\sqrt{d_n}, \dots, 1/\sqrt{d_n})^\top$, the following set of conditions relaxes the assumptions on the eigenstructure by strengthening the requirements on the coefficient matrices.

Lemma 5. *Suppose that*

- (i) $\sum_{k=1}^{\infty} |\pi_{nk}| = O(1)$ and
- (ii) $\sup_{n \geq 1} \sup_{\nu \geq 1} \sum_{j=1}^{d_n} |b_{n,k-r_\ell}^{(\nu,j)}| \stackrel{k,\ell}{\ll} (k+2r_\ell)^{-3-\theta}.$

Then $\sup_{n \geq 1} \max_{1 \leq \nu \leq d_n} |c_{nk}^{(Z,\nu)}| \stackrel{k}{\ll} k^{-3/4-\theta/2}$, i.e. (D) holds.

Proof of Lemma 5. Recall that $\|\mathbf{l}_{ni}\|_{\ell_2} = \|\mathbf{l}_{ni}\|_{\ell_2} = 1$ for all n, i . The proof is similar as the proof of Lemma 4 noting that $\|\mathbf{r}_i\|_{\ell_2}$ implies $|r_{i\ell}| \leq 1$ and using the estimate $|\mathbf{b}_{k-r_\ell,\nu}^{(\nu)}{}^\top \mathbf{l}_i r_{i\ell}| \leq \|\mathbf{b}_{k-r_\ell}\|_{\ell_2} \|\mathbf{l}_i\|_{\ell_2} |r_{i\ell}| \leq \sqrt{\sum_{\mu=1}^{d_n} (b_{k-r_\ell}^{(\nu,\mu)})^2} \stackrel{j,\ell}{\ll} (k+r_\ell)^{-3/2-\theta/2}.$ \square

Assumption (ii) is a weak localizing condition on the coefficient matrices of the multivariate linear process, as it limits the influence of the μ th innovation on the ν th coordinate process at all lags j . Observe that a sufficient condition for (ii) is to assume that

$$\sup_{n \geq 1} \sup_{1 \leq \nu, \mu} |b_{nj}^{(\nu,\mu)}| \stackrel{j,\ell}{\ll} (j+2r_\ell)^{-3/2-\theta/2} / \mu^{1/2+\delta}$$

for some $\delta > 0$.

Consider a stable VARMA model

$$(67) \quad \mathbf{Y}_{ni} = \mathbf{A}_{n1} \mathbf{Y}_{n,i-1} + \dots + \mathbf{A}_{np} \mathbf{Y}_{n,i-p} + \mathbf{M}_{n1} \boldsymbol{\epsilon}_{n,i-1} + \dots + \mathbf{M}_{nr} \boldsymbol{\epsilon}_{n,i-r} + \boldsymbol{\epsilon}_{ni},$$

as introduced in the main document with $(d_n \times d_n)$ coefficient matrices $\mathbf{A}_{n1}, \dots, \mathbf{A}_{np}$ and $\mathbf{M}_{n1}, \dots, \mathbf{M}_{nr}$ satisfying (element-wise) the decay condition (7) with $\varpi = 1$ for some $\delta > 0$. The coefficient matrices, $\boldsymbol{\Phi}_{nj}$, of the $MA(\infty)$ representation,

$$\mathbf{Y}_{ni} = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_{nj} \boldsymbol{\epsilon}_{n,i-j},$$

can be calculated using the recursion

$$\boldsymbol{\Phi}_{n0} = I_{d_n}, \boldsymbol{\Phi}_{nj} = \mathbf{M}_{nj} + \sum_{k=1}^j \mathbf{A}_{nk} \boldsymbol{\Phi}_{n,j-k}, j \geq 1,$$

where $\mathbf{M}_{nj} = \mathbf{0}$ for $j > q_n$. Denote $\mathbf{A}_{nk} = (a_{nk}^{(\nu,\mu)})_{\nu,\mu}$ and $\boldsymbol{\Phi}_j = (\Phi_{nj}^{(\nu,\mu)})_{\nu,\mu}$.

Lemma 6. *Suppose that the coefficient matrices of the VARMA model satisfy (7) with $\varpi = 1$, i.e. for some $\delta > 0$*

$$|a_{nk}^{(\nu,\mu)}|, |m_{nk}^{(\nu,\mu)}| \ll (k+2r_\ell)^{-5/2-\theta} (\nu\mu)^{-1/2-\delta}.$$

Then $|\Phi_{nj}^{(\nu,\mu)}| \stackrel{j,\ell}{\ll} (j+2r_\ell)^{-3/2-\theta} (\nu\mu)^{-1/2-\delta} \stackrel{\nu,\mu}{\ll} (j+2r_\ell)^{-3/2-\theta}$, such that (7) is satisfied with $\varpi = 0$, and therefore the decay condition (D) holds for the coefficients, $c_{nk}^{(Z,\nu)} = \sum_{\ell=1}^{q_n} \mathbf{1}(k \geq r_\ell) \Phi_{n,k-r_\ell}^{(\nu,\ell)}$, in the representation (4) of the multivariate linear process associated to the VARMA model (67).

Proof of Lemma 6. The proof is by induction. For $j = 0$ the assertion follows from $\Phi_{n0}^{(\nu,\mu)} = M_{n0}^{(\nu,\mu)} + A_{n0}^{(\nu,\mu)}$. For $j > 1$ it suffices to show that

$$\sum_{k=1}^j \sum_{s=1}^{d_n} |a_{nk}^{(\nu,s)} \Phi_{n,j-k}^{(s,\mu)}| \ll^{j,\ell} (j+2r_\ell)^{-3/2-\theta} (\nu\mu)^{-1/2-\delta}.$$

Since for $1 \leq k < j$

$$\begin{aligned} |a_{nk}^{(\nu,s)}| &\ll^{k,\ell} (k+2r_\ell)^{-5/2-\theta} (\nu s)^{-1/2-\delta}, \\ |\Phi_{n,j-k}^{(s,\mu)}| &\ll^{j,k,\ell} (j-k+2r_\ell)^{-3/2-\theta} (\mu s)^{-1/2-\delta}, \end{aligned}$$

we have

$$\begin{aligned} |a_{nk}^{(\nu,s)} \Phi_{n,j-k}^{(s,\mu)}| &\ll^{j,k,\ell} (k+2r_\ell)^{-5/2-\theta} (j-k+2r_\ell)^{-3/2-\theta} (s^2\nu\mu)^{-1/2-\delta} \\ &\ll^{j,k,\ell} [(k+2r_\ell)(j-k+2r_\ell)]^{-5/2-\theta} (s^2\nu\mu)^{-1/2-\delta}. \end{aligned}$$

Using the fact that $ab \geq a+b-1$ for $a, b \geq 1$, we obtain $(k+2r_\ell)(j-k+2r_\ell) \geq j+4r_\ell-1 \geq j+2r_\ell$, such that

$$|a_{nk}^{(\nu,s)} \Phi_{n,j-k}^{(s,\mu)}| \ll^{j,k,\ell} (j+2r_\ell)^{-5/2-\theta} (s^2\nu\mu)^{-1/2-\delta}.$$

Consequently,

$$\sum_{s=1}^{d_n} |a_{nk}^{(\nu,s)} \Phi_{n,j-k}^{(s,\mu)}| \ll^{j,k,\ell} (j+2r_\ell)^{-5/2-\theta} (\nu\mu)^{-1/2-\delta} \sum_{s=1}^{\infty} s^{-1-\delta/2}$$

and we may conclude that

$$\sum_{k=1}^j \sum_{s=1}^{d_n} |a_{nk}^{(\nu,s)} \Phi_{n,j-k}^{(s,\mu)}| \ll^{j,\ell} j(j+2r_\ell)^{-5/2-\theta} (\nu\mu)^{-1/2-\delta} \ll (j+2r_\ell)^{-3/2-\theta} (\nu\mu)^{-1/2-\delta},$$

such that

$$|\Phi_{nj}^{(\nu,\mu)}| \leq |M_{nj}^{(\nu,\mu)}| + \sum_{k=1}^j \sum_{s=1}^{d_n} |a_{nk}^{(\nu,s)} \Phi_{n,j-k}^{(s,\mu)}| \ll^{j,\ell} (j+2r_\ell)^{-3/2-\theta} (\nu\mu)^{-1/2-\delta} \ll^{\nu,\mu} (j+2r_\ell)^{-3/2-\theta}.$$

Using again the estimate (66), we may conclude that $|c_{nk}^{(Z,\nu)}| \leq \sum_{\ell=1}^{q_n} |\Phi_{n,k-r_\ell}^{(\nu,\ell)}| \ll^{\nu} k^{-3/4-\theta/2}$. \square

We need the following lemma.

Lemma 7. *Under Assumption (D) it holds for weighting vectors $\mathbf{v}_n, \mathbf{w}_n, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n$ with finite ℓ_1 -norms*

$$(68) \quad \sum_{i=1}^{\infty} \sum_{\ell=0}^{\infty} (\tilde{f}_{\ell,i}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) - \tilde{f}_{\ell,i}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n))^2 \ll^{n,n'} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{1-\theta} \quad \text{for } n', n = 1, 2, \dots,$$

$$(69) \quad \sum_{k=1}^n \sum_{r=0}^{\infty} [\tilde{f}_{r+k,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)]^2 \ll^{n,n'} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{1-\theta} \quad \text{for } n', n = 1, 2, \dots,$$

$$(70) \quad \sum_{k=1}^n \sum_{\ell=0}^{\infty} [\tilde{f}_{\ell,k}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)]^2 \ll^{n,n'} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{1-\theta} \quad \text{for } n', n = 1, 2, \dots.$$

The constants arising in the above estimates do not depend on \mathbf{a} .

Proof. Omitted for brevity. \square

The following lemma provides estimates needed to study a change of the coefficients.

Lemma 8. *Under Assumption (D) it holds for vectors $\mathbf{v}_n, \mathbf{w}_n$ with finite ℓ_1 -norms:*

$$(i) \quad |\tilde{f}_{\ell,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)| \ll^n \|\mathbf{v}_n\|_{\ell_1} \|\mathbf{w}_n\|_{\ell_1} \ell^{-3/4-\theta/2}, \quad \ell \geq 1.$$

$$(ii) \quad \sum_{j=1}^{n'} \sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n)]^2 \sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)]^2 \ll^{n,n'} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 \|\tilde{\mathbf{v}}_n\|_{\ell_1}^2 \|\tilde{\mathbf{w}}_n\|_{\ell_1}^2 (n')^{1-\theta}.$$

$$(iii) \sum_{j=1}^{n'} \sum_{\ell=j}^{\infty} \left| \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right| \stackrel{n,n'}{\ll} \|\mathbf{v}_n\|_{\ell_1} \|\mathbf{w}_n\|_{\ell_1} \|\tilde{\mathbf{v}}_n\|_{\ell_1} \|\tilde{\mathbf{w}}_n\|_{\ell_1} (n')^{1-\theta}.$$

Proof of Lemma 8. By virtue of Assumption (D), we have for $j, \ell \geq 1$ with $C_n = C_n(\mathbf{v}_n, \mathbf{w}_n) = \|\mathbf{v}_n\|_{\ell_1} \|\mathbf{w}_n\|_{\ell_1}$

$$\begin{aligned} |f_{l,j}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)| &\stackrel{n}{\ll} C_n j^{-3/4-\theta/2} (\ell+j)^{-3/4-\theta/2} \\ &\stackrel{n}{\ll} C_n j^{-1-\theta/4} j^{1/4-\theta/4} (\ell+j)^{-3/4-\theta/2} \\ &\stackrel{n}{\ll} C_n j^{-1-\theta/4} (\ell+j)^{-1/2-\frac{3}{4}\theta} \\ &\stackrel{n}{\ll} C_n j^{-1-\theta/4} \ell^{-1/2-\frac{3}{4}\theta}. \end{aligned}$$

Hence $\sum_{j=1}^{\infty} |f_{\ell,j}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)| \stackrel{n}{\ll} C_n \ell^{-1/2-\frac{3}{4}\theta}$. Since $|f_{\ell,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)| \stackrel{n}{\ll} C_n \ell^{-3/4-\theta/2}$ for $\ell \geq 1$, we obtain

$$|\tilde{f}_{\ell,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)| \stackrel{n}{\ll} C_n \ell^{-1/2-\frac{3}{4}\theta}, \quad \ell \geq 1.$$

Next we show (ii). Put $K_n = C_n(\mathbf{v}_n, \mathbf{w}_n)C_n(\tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$. By (i) we have for $j \geq 1$

$$\sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)]^2 \stackrel{n}{\ll} C_n^2 \sum_{\ell=j}^{\infty} \ell^{-1-\frac{3}{2}\theta} \stackrel{n}{\ll} C_n^2 \lim_{b \rightarrow \infty} \int_j^b x^{-1-\frac{3}{2}\theta} dx \stackrel{n}{\ll} C_n^2 j^{-\frac{3}{2}\theta},$$

such that

$$\begin{aligned} \frac{1}{n'} \sum_{j=1}^{n'} \sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n)]^2 \sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)]^2 &\stackrel{n,n'}{\ll} K_n^2 (n')^{-\theta} \sum_{j=1}^{n'} (1/j)^{3\theta} (1/n')^{1-\theta} \\ &\stackrel{n,n'}{\ll} K_n^2 (n')^{-\theta} \sum_{j=1}^{n'} j^{-1-4\theta} \stackrel{n,n'}{\ll} K_n^2 (n')^{-\theta}, \end{aligned}$$

which establishes (ii). To show (iii), observe that for $j \geq 1$

$$\sum_{\ell=j}^{\infty} \left| \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right| \stackrel{n}{\ll} K_n \sum_{\ell=j}^{\infty} \ell^{-1-\frac{3}{2}\theta} \stackrel{n}{\ll} K_n \lim_{b \rightarrow \infty} \int_j^b x^{-1-\frac{3}{2}\theta} dx \stackrel{n}{\ll} K_n j^{-\frac{3}{2}\theta}.$$

Therefore, for $n' \geq 1$

$$\begin{aligned} \frac{1}{n'} \sum_{j=1}^{n'} \sum_{\ell=j}^{\infty} \left| \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right| &\stackrel{n,n'}{\ll} K_n \frac{1}{n'} \sum_{j=1}^{n'} j^{-\frac{3}{2}\theta} \\ &\stackrel{n,n'}{\ll} K_n (n')^{-\theta} \sum_{j=1}^{n'} (1/j)^{\frac{3}{2}\theta} (1/n')^{1-\theta} \\ &\stackrel{n,n'}{\ll} K_n (n')^{-\theta} \sum_{j=1}^{\infty} j^{-1-\theta/2} \stackrel{n,n'}{\ll} K_n (n')^{-\theta}. \end{aligned}$$

□

A.2. Proofs for Subsection 7.2 (Martingale Approximations). Recall the following definitions: Introduce for coefficients \mathbf{a} satisfying Assumption (D) and vectors \mathbf{v}_n and \mathbf{w}_n the \mathcal{F}_{nk} -martingales

$$M_k^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = \tilde{f}_{0,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) \sum_{i=0}^k (\epsilon_{ni}^2 - \sigma_i^2) + \sum_{i=0}^k \epsilon_{ni} \sum_{l=1}^{\infty} \tilde{f}_{l,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) \epsilon_{n,i-j}, \quad k \geq 0,$$

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

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

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

$$(71) \quad S_{k,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) = \mathbf{D}_{nk}^{(1)}, \quad S_{k,0}^{(n)}(\mathbf{c}, \mathbf{v}_n, \mathbf{w}_n) = \mathbf{D}_{nk}^{(2)},$$

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

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

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

$$(72) \quad \delta M_k^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = \tilde{f}_{0,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) \sum_{i=1}^k (\epsilon_{ni}^2 - \sigma_{ni}^2) + \sum_{i=1}^k \epsilon_{ni} \sum_{l=1}^{\infty} \tilde{f}_{l,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) \epsilon_{n,i-l}, \quad k, n \geq 1,$$

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

$$(73) \quad \begin{aligned} &\mathbb{C}(\delta M_{m'+n'}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n), \delta M_{m'+n'}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)) \\ &= \tilde{f}_{0,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{0,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \sum_{j=1}^{n'} (\gamma_{n,m'+j} + \sigma_{n,m'+j}^4) + \sum_{j=1}^{n'} \sum_{\ell=1}^{\infty} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \sigma_{n,m'+j}^2 \sigma_{n,m'+j-\ell}^2, \end{aligned}$$

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

Proof of Lemma 1. For $n, n' \geq 1$ and $m' \geq 0$ put

$$\begin{aligned} \beta_{n,n',m'}^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) &= \tilde{f}_{0,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{0,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \frac{1}{n'} \sum_{j=1}^{n'} (\gamma_{n,m'+j} - \sigma_{n,m'+j}^4) \\ &\quad + \frac{1}{n'} \sum_{j=1}^{n'} \sigma_{n,m'+j}^2 \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \sigma_{n,m'+j-\ell}^2. \end{aligned}$$

Then

$$\begin{aligned} &\left| \mathbb{C}(\delta M_{m'+n'}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n), \delta M_{m'+n'}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)) - (n') \beta_{n,n',m'}^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right| \\ &\leq \left(\sup_k E(\epsilon_k^2) \right)^2 \sum_{j=1}^{n'} \sum_{\ell=1}^{\infty} \left| \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right| \\ &\quad \ll_{n,n',m'} K_n(n')^{1-\theta}, \end{aligned}$$

by Lemma 8 (iii). We shall prove that one may replace $\beta_{n,n',m'}^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$ by

$$\beta_{n,n'}^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) = \tilde{f}_{0,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{0,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) (\gamma_n - s_{n1}^4) + \frac{s_{n1}^4}{n'} \sum_{j=1}^{n'} \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$$

with an error term of order $O(K_n \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{1-\theta})$. Then a further application of Lemma 8 (iii) shows that the range of summation for ℓ can be extended to \mathbb{N} and gives $|\beta_{n,n'}^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) -$

$\beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)| \ll_{n,n'} K_n \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{1-\theta}$, such that the assertion follows. First observe the

following fact: If $\{\alpha_n^*, \alpha_{nk} : k \geq 1\} \subset \mathbb{R}$, $n \geq 1$, satisfy $(n')^{-1} \sum_{i=1}^{n'} i |\alpha_{ni} - \alpha_n^*| \ll_{n,n'} (n')^{-\beta}$ for some $\beta > 1$, then

$$\sup_{m' \geq 1} \frac{1}{n'} \sum_{i=m'+1}^{m'+n'} i |\alpha_{ni} - \alpha_n^*| \ll_{n,n',m'} (n')^{-1} \ll_{n,n',m'} (n')^{-\theta}$$

This follows from

$$\frac{1}{n'} \sum_{i=m'+1}^{m'+n'} i |\alpha_{ni} - \alpha_n^*| = \frac{n' + m'}{n'} \left\{ \frac{1}{n' + m'} \sum_{i=1}^{m'+n'} i |\alpha_{ni} - \alpha_n^*| - \frac{m'}{m' + n'} \frac{1}{m'} \sum_{i=1}^{m'} i |\alpha_{ni} - \alpha_n^*| \right\}$$

which implies

$$\begin{aligned} \frac{1}{n'} \sum_{i=m'+1}^{m'+n'} i |\alpha_{ni} - \alpha_n^*| &\stackrel{n, n', m'}{\ll} \frac{n' + m'}{n'} (m' + n')^{-\beta} + \frac{m'}{n'} (m')^{-\beta} \\ &\stackrel{n, n', m'}{\ll} (n')^{-1} (m' + n')^{1-\beta} + (n')^{-1} (m')^{1-\beta} \stackrel{n, n', m'}{\ll} (n')^{-1}, \end{aligned}$$

since $\beta > 1$. Next observe that

$$(74) \quad \tilde{f}_{0,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n), \max_{j=1, \dots, n'} \max_{\ell=1, \dots, n'-1} \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \stackrel{n,n}{\ll} \|\mathbf{v}_n\|_{\ell_1} \|\mathbf{w}_n\|_{\ell_1},$$

and analogous estimates hold for the triple $(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$. Because $\theta \leq \beta$ we have $(n')^{-\beta} \leq (n')^{-\theta}$ and therefore by using

$$(75) \quad \frac{1}{n'} \sum_{i=1}^{n'} i |\sigma_{ni}^2 - s_{n1}^2| \stackrel{n, n'}{\ll} (n')^{-\beta}$$

and

$$(76) \quad \frac{1}{n'} \sum_{i=1}^{n'} i |\gamma_{ni} - \gamma_n| \stackrel{n, n'}{\ll} (n')^{-\beta}$$

for constants $s_{n1}^2 \in (0, \infty)$ and $\gamma_n \in \mathbb{R}$ for some $1 < \beta < 2$ with $1 + \theta < \beta$, and the decomposition

$$\sigma_{n,m'+j}^4 - s_{n1}^4 = (\sigma_{n,m'+j}^2 - s_{n1}^2)(\sigma_{n,m'+j}^2 + s_{n1}^2)$$

where $(n')^{-1} \sum_{j=1}^{n'} \sigma_{n,m'+j}^2 \ll 1$, we obtain

$$(77) \quad \left| \tilde{f}_{0,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{0,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \frac{1}{n'} \sum_{j=1}^{n'} (\gamma_{n,m'+j} - \sigma_{n,m'+j}^4) - \tilde{f}_{0,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{0,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) (\gamma_n - s_{n1}^4) \right| \stackrel{n, n'}{\ll} K_n (n')^{-\theta}.$$

Next we show that the second term of $\beta_{n,n',m'}^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$ can be replaced by the second term of $\beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)$. Use

$$\sigma_{n,m'+j}^2 \sigma_{n,m'+\ell}^2 - s_{n1}^4 = (\sigma_{n,m'+j}^2 - s_{n1}^2) \sigma_{n,m'+\ell}^2 + (\sigma_{n,m'+\ell}^2 - s_{n1}^2) s_{n1}^2$$

to obtain

$$\begin{aligned} &\frac{1}{n'} \sum_{j=1}^{n'} \sigma_{n,m'+j}^2 \sum_{\ell=1}^{j-1} \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \sigma_{n,m'+j-\ell}^2 - \frac{s_{n1}^4}{n'} \sum_{j=1}^{n'} \sum_{\ell=1}^{j-1} \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \\ &= \frac{1}{n'} \sum_{j=1}^{n'} \sum_{\ell=1}^{j-1} \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) [\sigma_{n,m'+j}^2 \sigma_{n,m'+\ell}^2 - s_{n1}^4] \\ &= \frac{1}{n'} \sum_{j=1}^{n'} \sum_{\ell=1}^{j-1} \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) (\sigma_{n,m'+j}^2 - s_{n1}^2) \sigma_{n,m'+\ell}^2 \\ &\quad + \frac{1}{n'} \sum_{j=1}^{n'} \sum_{\ell=1}^{j-1} \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) (\sigma_{n,m'+\ell}^2 - s_{n1}^2) s_{n1}^2. \end{aligned}$$

The first term of the last decomposition can be estimated as follows.

$$\begin{aligned}
& \left| \frac{1}{n'} \sum_{j=1}^{n'} \sum_{\ell=1}^{j-1} \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) (\sigma_{n,m'+j}^2 - s_{n1}^2) \sigma_{n,m'+\ell}^2 \right| \\
& \stackrel{n,n',m'}{\ll} \frac{K_n}{n'} \sum_{j=1}^{n'} \sum_{\ell=1}^{j-1} |\sigma_{n,m'+j}^2 - s_{n1}^2| \sigma_{n,m'+\ell}^2 \\
& \leq \frac{K_n}{n'} \sum_{j=2}^{n'} (j-1) |\sigma_{n,m'+j}^2 - s_{n1}^2| \frac{1}{j-1} \sum_{\ell=1}^{j-1} |\sigma_{n,m'+\ell}^2 - s_{n1}^2| + \frac{K_n}{n'} \sum_{j=2}^{n'} s_{n1}^2 (j-1) |\sigma_{n,m'+j}^2 - s_{n1}^2| \\
& \stackrel{n,n',m'}{\ll} \frac{K_n}{n'} \sum_{j=2}^{n'} |\sigma_{n,m'+j}^2 + s_{n1}^2| (j-1)^{1-\beta} + K_n (n')^{-\beta} \stackrel{n',m'}{\ll} K_n (n')^{-\beta} \stackrel{n'}{\ll} K_n (n')^{-\theta}.
\end{aligned}$$

Similarly, for the second term we have

$$\begin{aligned}
& \left| \frac{1}{n'} \sum_{j=1}^{n'} \sum_{\ell=1}^{j-1} \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{j-\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) (\sigma_{n,m'+\ell}^2 - s_{n1}^2) s_{n1}^2 \right| \\
& \stackrel{n,n',m'}{\ll} \frac{K_n}{n'} \sum_{j=2}^{n'} \sum_{\ell=1}^{j-1} |\sigma_{n,m'+\ell}^2 - s_{n1}^2| \stackrel{n,n'}{\ll} \frac{K_n}{n'} \sum_{j=2}^{n'} (j-1)^{1-\beta} \stackrel{n'}{\ll} K_n (n')^{-\theta'} \stackrel{n'}{\ll} K_n (n')^{-\theta}
\end{aligned}$$

for $\theta < \theta' < \beta - 1$ (θ' exists, since $\theta < \beta - 1$ by assumption). Putting things together, we arrive at

$$\left| \frac{1}{n'} \sum_{j=1}^{n'} \sigma_{n,m'+j}^2 \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \sigma_{n,m'+j-\ell}^2 - \frac{s_1^4}{n'} \sum_{j=1}^{n'} \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right| \ll K_n (n')^{-\theta}.$$

Combining the latter estimate with (77) shows that

$$|\beta_{n,n',m'}^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) - \beta_{n,n'}^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)| \stackrel{n,n',m'}{\ll} K_n (n')^{-\theta},$$

which completes the proof. \square

Proof of Lemma 2. For brevity of notation, we omit the dependence on $\mathbf{v}_n, \mathbf{w}_n$ in notation. Put $R_{n',m'}^{(n)}(\mathbf{a}) = S_{n',m'}^{(n)}(\mathbf{a}) - \delta M_{m'+n'}^{(n)}(\mathbf{a})$ and note that $R_{n'}^{(n)}(\mathbf{a}) = O_{n',m'}^{(n)}(\mathbf{a}) + P_{n',m'}^{(n)}(\mathbf{a}) + Q_{n',m'}^{(n)}(\mathbf{a})$, where

$$\begin{aligned}
Q_{n',m'}^{(n)}(\mathbf{a}) &= \sum_{i=1}^{n'-1} \sum_{\ell=0}^{n'-i-1} \tilde{f}_{\ell,i+1}^{(n)}(\mathbf{a}) (\sigma_{n,m'+n'-i}^2 \mathbf{1}(\ell=0) - \epsilon_{n,m'+n'-i} \epsilon_{n,m'+n'-i-\ell}), \\
P_{n',m'}^{(n)}(\mathbf{a}) &= \sum_{i=0}^{\infty} \sum_{\ell=0}^{\infty} (\tilde{f}_{\ell,i+1}^{(n)}(\mathbf{a}) - \tilde{f}_{\ell,i+n'+1}^{(n)}(\mathbf{a})) (\epsilon_{n,m'-i} \epsilon_{n,m'-i-\ell} - \sigma_{n,m'-i}^2 \mathbf{1}(\ell=0)), \\
O_{n',m'}^{(n)}(\mathbf{a}) &= - \sum_{i=0}^{n'-1} \sum_{k=n'}^{\infty} \tilde{f}_{k-i,i+1}^{(n)}(\mathbf{a}) \epsilon_{n,m'+n'-k} \epsilon_{n,m'+n'-i},
\end{aligned}$$

for $n', m' \geq 0$. The result can now be shown along the lines of [Kouritzin, 1995, Lemma 2] noting the following facts. By independence of $\{\epsilon_{nk} : k \in \mathbb{Z}\}$, for any fixed n ,

$$\mathbb{E}(Q_{n',m'}^{(n)})^2 = \sum_{i=0}^{n'-1} [\tilde{f}_{0,i+1}^{(n)}(\mathbf{a})]^2 (\gamma_{n,m'+n'-i} - \sigma_{n,m'+n'-i}^4) + \sum_{i=0}^{n'-1} \sum_{\ell=1}^{n'-i-1} [\tilde{f}_{\ell,i+1}^{(n)}(\mathbf{a})]^2 \sigma_{n,m'-i}^2 \sigma_{n,m'-i-\ell}^2.$$

By virtue of (68) - (70), which hold due to the decay assumption (D) on the coefficients of the vector time series, we obtain

$$\mathbb{E}(Q_{n',m'}^{(n)})^2 \stackrel{n,n',m'}{\ll} \sum_{i=0}^{n'-1} [\tilde{f}_{0,i+1}^{(n)}(\mathbf{a})]^2 + \sum_{i=0}^{n'-1} \sum_{\ell=1}^{n'-i-1} [\tilde{f}_{\ell,i+1}^{(n)}(\mathbf{a})]^2 \stackrel{n,n',m'}{\ll} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{1-\theta},$$

since by assumption $\sup_{n \geq 1} \sup_{k \geq 1} |\gamma_{n,k}| < \infty$ and $\sup_{n \geq 1} \sup_{k \geq 1} \mathbb{E}|\epsilon_{n,k}|^2 < \infty$, which entails that the rate $(n')^{1-\theta}$ also applies to innovation arrays $\{\epsilon_{nk} : k \in \mathbb{Z}, n \in \mathbb{N}\}$ satisfying Assumption (E). Analogously,

$$\mathbb{E}(P_{n',m'}^{(n)})^2 \stackrel{n,n',m'}{\ll} \sum_{i=1}^{\infty} \sum_{\ell=0}^{\infty} \left(\tilde{f}_{\ell,i}^{(n)}(\mathbf{a}) - \tilde{f}_{\ell,i+n'}^{(n)}(\mathbf{a}) \right)^2 \stackrel{n,n',m'}{\ll} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{1-\theta}.$$

Lastly,

$$\mathbb{E}(O_{n',m'}^{(n)}(\mathbf{a}))^2 = \mathbb{E} \lim_{N \rightarrow \infty} \left(\sum_{i=0}^{n'-1} \sum_{k=n'}^N \tilde{f}_{k-i,i+1}^{(n)}(\mathbf{a}) \epsilon_{n,m'+n'-k} \epsilon_{n,m'+n'-i} \right)^2,$$

and therefore Fatou's lemma leads to $\mathbb{E}(O_{n',m'}^{(n)}(\mathbf{a}))^2 \stackrel{n,n',m'}{\ll} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{1-\theta}$. Repeating the arguments provided in [Kouritzin, 1995] shows that $R_{n'}$ can be decomposed in three terms which are bounded by the expressions listed in (68) - (70), which are $\stackrel{n',n}{\ll} (n')^{1-\theta}$. Observing that the dependence on the vectors $\mathbf{v}_n, \mathbf{w}_n, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n$ is only through the coefficients $\tilde{f}_{\ell,j}(\cdot)$, the remaining assertions follow by recalling (71). \square

Observe that for i.i.d. error terms with $\mathbb{E}(\epsilon_k^2) = \sigma^2$ and $\mathbb{E}(\epsilon_k^3) = \gamma$ for all k

$$\beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) = \tilde{f}_{0,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{0,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) (\gamma + \sigma^4) + \sigma^4 \sum_{\ell=1}^{\infty} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n).$$

We write $\alpha_n^2(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = \beta_n^2(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)$, $\mathbf{a} \in \{\mathbf{b}, \mathbf{c}\}$. If $c_{nj}^{(\nu)} = c_j^{(\nu)}$, $n, \nu \geq 1$, then for projections $\mathbf{v}, \mathbf{w} \in \ell_1$ these quantities do not depend on n .

Proof of Lemma 3. We may apply the method of proof of [Kouritzin, 1995]. Define for a sequence \mathbf{a} satisfying Assumption (D) the following approximation for $M_{m'+j}^{(n)}(\mathbf{a}) - M_{m'}^{(n)}(\mathbf{a})$,

$$L_j(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = L_{n,m',j}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = \sum_{k=m'+1}^{m'+j} \{ \tilde{f}_{0,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) (\epsilon_{nk}^2 - \sigma_{nk}^2) + \sum_{\ell=1}^{k-m'-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) \epsilon_{nk} \epsilon_{n,k-\ell} \}$$

and let

$$K_j(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = M_{m'+j}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) - M_{m'}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) - L_j(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)$$

denote the associated approximation error, for $j > 0$, and $K_j(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = 0$, if $j = 0$. Then the lag 1 martingale differences attain the representation

$$\Delta M_{m'+j}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) := M_{m'+j}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) - M_{m'+j-1}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = \Delta L_j(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) + \Delta K_j(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n).$$

where

$$(79) \quad \Delta L_j(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = \tilde{f}_{0,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) (\epsilon_{n,m'+j}^2 - \sigma_{n,m'+j}^2) + \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) \epsilon_{n,m'+j} \epsilon_{n,m'+j-\ell}.$$

Using the fact that the lag n' difference operator is the sum of first order differences, i.e., $\delta M_{m'+n'}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = M_{m'+n'}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) - M_{m'}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n) = \sum_{j=1}^{n'} \Delta M_{m'+j}^{(n)}(\mathbf{a}, \mathbf{v}_n, \mathbf{w}_n)$, we obtain

$$\begin{aligned} E_{n'}^{(n)} &= \left\| \sum_{j=1}^{n'} \left\{ \mathbb{E}[\delta M_{m'+j}^{(n)}(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \delta M_{m'+j}(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) | \mathcal{F}_{n,m'}] - \beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right\} \right\|_{L_1} \\ &\leq \left\| \sum_{j=1}^{n'} \left\{ \mathbb{E}[\Delta L_j(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \Delta L_j(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) | \mathcal{F}_{n,m'}] - \beta_n^2(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n, \mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) \right\} \right\|_{L_1} \\ &\quad + \left\| \sum_{j=1}^{n'} \mathbb{E}[\Delta K_j(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \Delta L_j(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) | \mathcal{F}_{n,m'}] \right\|_{L_1} + \left\| \sum_{j=1}^{n'} \mathbb{E}[\Delta L_j(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \Delta K_j(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) | \mathcal{F}_{n,m'}] \right\|_{L_1} \\ &\quad + \left\| \sum_{j=1}^{n'} \mathbb{E}[\Delta K_j(\mathbf{b}, \mathbf{v}_n, \mathbf{w}_n) \Delta K_j(\mathbf{c}, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n) | \mathcal{F}_{n,m'}] \right\|_{L_1}. \end{aligned}$$

Let us now estimate the four terms separately. For brevity of notation we omit the dependence on $\mathbf{v}_n, \mathbf{w}_n, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n$, as they are attached to \mathbf{b} and \mathbf{c} , respectively, and enter only through the coefficients $\tilde{f}_{\ell,j}^{(n)}$. Using (79) we have

$$\begin{aligned} \Delta L_j(\mathbf{b}) \Delta L_j(\mathbf{c}) &= \tilde{f}_{0,0}^{(n)}(\mathbf{b}) \tilde{f}_{0,0}^{(n)}(\mathbf{c}) (\epsilon_{n,m'+j}^2 - \sigma_{n,m'+j}^2)^2 + \tilde{f}_{0,0}^{(n)}(\mathbf{b}) (\epsilon_{n,m'+j}^2 - \sigma_{n,m'+j}^2) \epsilon_{n,m'}^2 \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}) \epsilon_{n,m'+j-\ell} \\ &\quad + \tilde{f}_{0,0}^{(n)}(\mathbf{c}) (\epsilon_{n,m'+j}^2 - \sigma_{n,m'+j}^2) \epsilon_{n,m'}^2 \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}) \epsilon_{n,m'+j-\ell} + \sum_{\ell,\ell'=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}) \tilde{f}_{\ell',0}^{(n)}(\mathbf{c}) \epsilon_{n,m'}^2 \epsilon_{n,m'+j-\ell} \epsilon_{n,m'+j-\ell'} \end{aligned}$$

Noting that $j \geq 1$ and the sums over ℓ, ℓ' are non-vanishing only if $j \geq 2$, we obtain by independence of $\{\epsilon_{nk} : k \in \mathbb{Z}\}$ for $j \geq 2$ with the centered r.v.s. $\bar{\epsilon}_{n,m'+j}^2 = \epsilon_{n,m'+j}^2 - \sigma_{n,m'+j}^2$

$$\mathbb{E} \left[\tilde{f}_{0,0}^{(n)}(\mathbf{b}) \bar{\epsilon}_{n,m'+j}^2 \epsilon_{n,m'}^2 \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}) \epsilon_{n,m'+j-\ell} | \mathcal{F}_{n,m'} \right] = \tilde{f}_{0,0}^{(n)}(\mathbf{b}) \epsilon_{n,m'}^2 \mathbb{E}(\bar{\epsilon}_{n,m'+j}^2) \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}) \mathbb{E}(\epsilon_{n,m'+j-\ell}) = 0,$$

a.s., since $\epsilon_{n,m'+j}$ and $\epsilon_{n,m'+j-\ell}$ are independent if $j \geq 2$ and $\ell \geq 1$. Therefore

$$\mathbb{E}[\Delta L_j(\mathbf{b}) \Delta L_j(\mathbf{c}) | \mathcal{F}_{n,m'}] = \tilde{f}_{0,0}^{(n)}(\mathbf{b}) \tilde{f}_{0,0}^{(n)}(\mathbf{c}) (\gamma_{n,m'+j} - \sigma_{n,m'+j}^4) + \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}) \sigma_{n,m'+j}^2 \sigma_{n,m'+j-\ell}^2,$$

a.s.. Consequently, cf. (73),

$$\begin{aligned} &\sum_{j=1}^{n'} \{ \mathbb{E}[\Delta L_j(\mathbf{b}) \Delta L_j(\mathbf{c}) | \mathcal{F}_{n,m'}] - \beta_n^2(\mathbf{b}, \mathbf{c}) \} \\ &= \tilde{f}_{0,0}^{(n)}(\mathbf{b}) \tilde{f}_{0,0}^{(n)}(\mathbf{c}) \sum_{j=1}^{n'} (\gamma_{n,m'+j} - \sigma_{n,m'+j}^4) + \sum_{j=1}^{n'} \sigma_{n,m'+j}^2 \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}) \sigma_{n,m'+j-\ell}^2 - (n') \beta_n^2(\mathbf{b}, \mathbf{c}), \\ &\ll K_n(n')^{1-\theta}. \end{aligned}$$

a.s.. A lengthy calculation shows that $\Delta K_j(\mathbf{a}) = \epsilon_{n,m'+j} \sum_{\ell=j}^{\infty} \tilde{f}_{\ell,0}^{(n)}(\mathbf{a}) \epsilon_{n,m'+j-\ell}$, such that, because $\epsilon_{n,m'+j-\ell} \epsilon_{n,m'+j-\ell'}$ is $\mathcal{F}_{n,m'}$ -measurable if $\ell, \ell' \geq j$ and $j \geq 1$,

$$\mathbb{E}[\Delta K_j(\mathbf{b}) \Delta K_j(\mathbf{c}) | \mathcal{F}_{n,m'}] = \sigma_{n,m'+j}^2 \left(\sum_{\ell=j}^{\infty} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}) \epsilon_{n,m'+j-\ell} \right) \left(\sum_{\ell=j}^{\infty} \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}) \epsilon_{n,m'+j-\ell} \right),$$

such that the Cauchy-Schwarz inequality provides us with the bound

$$\|\mathbb{E}[\Delta K_j(\mathbf{b})\Delta K_j(\mathbf{c}) \mid \mathcal{F}_{n,m'}]\|_{L_1} \leq \left(\sup_{n,k \geq 1} \sigma_{nk}^2 \right)^3 \prod_{\mathbf{a} \in \{\mathbf{b}, \mathbf{c}\}} \sqrt{\sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{a})]^2},$$

a.s.. This implies

$$\begin{aligned} \left\| \sum_{j=1}^{n'} \mathbb{E}[\Delta K_j(\mathbf{b})\Delta K_j(\mathbf{c}) \mid \mathcal{F}_{n,m'}] \right\|_{L_1} &\leq \sum_{j=1}^{n'} \|\mathbb{E}[\Delta K_j(\mathbf{b})\Delta K_j(\mathbf{c}) \mid \mathcal{F}_{n,m'}]\|_{L_1} \\ &\ll^{n,n'} \sum_{j=1}^{n'} \prod_{\mathbf{a} \in \{\mathbf{b}, \mathbf{c}\}} \sqrt{\sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{a})]^2} \ll^{n,n'} K_n(n')^{1-\theta/2}, \end{aligned}$$

since by virtue of the Jensen inequality and Lemma 8

$$\left(\frac{1}{n'} \sum_{j=1}^{n'} \sqrt{\prod_{\mathbf{a} \in \{\mathbf{b}, \mathbf{c}\}} \sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{a})]^2} \right)^2 \leq \frac{1}{n'} \sum_{j=1}^{n'} \prod_{\mathbf{a} \in \{\mathbf{b}, \mathbf{c}\}} \sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{a})]^2.$$

Further,

$$\mathbb{E}[\Delta L_j(\mathbf{b})\Delta K_j(\mathbf{b}) \mid \mathcal{F}_{n,m'}] = \tilde{f}_{0,0}^{(n)}(\mathbf{b}) E(\epsilon_{nk}^3) \sum_{\ell=j}^{\infty} \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}) \epsilon_{n,m'+j-\ell}$$

leading to the estimate

$$\|\mathbb{E}[\Delta L_j(\mathbf{b})\Delta K_j(\mathbf{c}) \mid \mathcal{F}_{n,m'}]\|_{L_1} \leq \sup_{n,k} \mathbb{E}|\epsilon_{nk}|^3 \sup_{n,k} \mathbb{E}(\epsilon_{nk}^2) \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}) \sqrt{\sum_{\ell=j}^{\infty} [\tilde{f}_{\ell,0}^{(n)}(\mathbf{c})]^2} \ll^{n,n'} K_n(n')^{1-\theta/2}.$$

Lastly, a direct calculation using similar arguments as above shows that

$$\mathbb{E}[\Delta_j(\mathbf{b})\Delta L_j(\mathbf{c}) \mid \mathcal{F}_{n,m'}] = \tilde{f}_{0,0}^{(n)}(\mathbf{b}) \tilde{f}_{0,0}^{(n)}(\mathbf{c}) [\mathbb{E}(\epsilon_{n,m'+j}^4) - \sigma_{n,m'+j-\ell}^4] + \sum_{\ell=1}^{j-1} \tilde{f}_{\ell,0}^{(n)}(\mathbf{b}) \tilde{f}_{\ell,0}^{(n)}(\mathbf{c}) \sigma_{n,m'+j}^2 \sigma_{n,m'+j-\ell}^2,$$

which is the j th term of the first sum of $(n')\beta_{n,n',m'}^2(\mathbf{b})$ as defined in the proof of Lemma 1. Since

there it was shown that $|\beta_{n,n',m'}^2(\mathbf{b}) - \beta_n^2(\mathbf{b})| \ll^{n,n'} \|\mathbf{v}_n\|_{\ell_1}^2 \|\mathbf{w}_n\|_{\ell_1}^2 (n')^{-\theta}$, we eventually obtain

$$\sum_{j=1}^{n'} \mathbb{E}[\Delta_j(\mathbf{b})\Delta L_j(\mathbf{c}) \mid \mathcal{F}_{n,m'}] - (n')\beta_n^2(\mathbf{b}, \mathbf{c}) \ll K_n(n')^{1-\theta/2}.$$

Putting together the above estimates completes the proof of the first assertion. The second assertion is shown as in [Kouritzin, 1995, (4.23)] and is omitted for brevity. \square

A.3. Proofs of Subsection 7.3.

Lemma 9. *Under the change-point model (9) it holds*

$$(80) \quad \mathbb{E} \left(\mathbf{S}_{nk} - \frac{k}{n} \mathbf{S}_{nn} \right) = \begin{cases} \frac{k(n-\tau)}{n} (\boldsymbol{\Sigma}_{n0} - \boldsymbol{\Sigma}_{n1}), & k \leq \tau, \\ \tau \frac{n-k}{n} (\boldsymbol{\Sigma}_{n0} - \boldsymbol{\Sigma}_{n1}), & k > \tau. \end{cases}$$

and

$$(81) \quad m_n(k) := \mathbb{E} \left(U_{nk} - \frac{k}{n} U_{nn} \right) = \begin{cases} \frac{k(n-\tau)}{n} \Delta_n, & k \leq \tau, \\ \tau \frac{n-k}{n} \Delta_n, & k > \tau. \end{cases}$$

Proof of Lemma 9. Since $\mathbb{E}(\mathbf{Y}_{ni} \mathbf{Y}_{ni}^\top) = \boldsymbol{\Sigma}_{n0} \mathbf{1}(i \leq \tau) + \boldsymbol{\Sigma}_{n1} \mathbf{1}(i > \tau)$, we have $\mathbb{E}(\frac{1}{n} \mathbf{S}_{nn}) = \mathbb{E}(\frac{1}{n} \sum_{i=1}^n \mathbf{Y}_{ni} \mathbf{Y}_{ni}^\top) = \frac{\tau}{n} \boldsymbol{\Sigma}_{n0} + \frac{n-\tau}{n} \boldsymbol{\Sigma}_{n1}$. Therefore, for $k \leq \tau$ $\mathbb{E}(\mathbf{S}_{nk} - \frac{k}{n} \mathbf{S}_{nn}) = k \boldsymbol{\Sigma}_{n0} - k (\frac{\tau}{n} \boldsymbol{\Sigma}_{n0} + \frac{n-\tau}{n} \boldsymbol{\Sigma}_{n1}) = \frac{k(n-\tau)}{n} (\boldsymbol{\Sigma}_{n0} - \boldsymbol{\Sigma}_{n1})$, whereas for $k > \tau$ $\mathbb{E}(\mathbf{S}_{nk} - \frac{k}{n} \mathbf{S}_{nn}) = \tau \boldsymbol{\Sigma}_{n0} + (\tau - k) \boldsymbol{\Sigma}_{n1} - k (\frac{\tau}{n} \boldsymbol{\Sigma}_{n0} + \frac{n-\tau}{n} \boldsymbol{\Sigma}_{n1}) = \tau \frac{n-k}{n} (\boldsymbol{\Sigma}_{n0} - \boldsymbol{\Sigma}_{n1})$. This verifies (80). Recalling that $U_{nk} = \mathbf{v}_n^\top \mathbf{S}_{nk} \mathbf{w}_n$ and $\Delta_n = \mathbf{v}_n^\top \boldsymbol{\Sigma}_{n0} \mathbf{w}_n - \mathbf{v}_n^\top \boldsymbol{\Sigma}_{n1} \mathbf{w}_n$, (81) follows by linearity. \square

Proof of Theorem 2. Observe that $\max_{k \leq n} |U_{nk} - \frac{k}{n} U_{nn}| = \max_{k \leq n} \frac{1}{\sqrt{n}} |D_{nk} - \frac{k}{n} D_{nn} + m_n(k)|$.

In Theorem 11 it is shown that \tilde{D}_{nk} , defined in (85), is a martingale which approximates D_{nk} , since $\mathbb{E}(\tilde{D}_{nk} - D_{nk})^2 \stackrel{n,k}{\ll} n^{1-\theta}$, so that $\mathbb{E}(\tilde{D}_{nk})^2 \stackrel{n,k}{\ll} k$ as well as $\mathbb{E}(\tilde{D}_{nk} - \frac{k}{n} \tilde{D}_{nn})^2 \stackrel{n,k}{\ll} n$ hold. Using this fact, (83) and the triangle inequality, we obtain for any constant $C > 0$ and $k \leq n$

$$\begin{aligned} \mathbb{P}\left(\frac{1}{\sqrt{n}} \left|D_{nk} - \frac{k}{n} D_{nn}\right| > C\right) &\leq \mathbb{P}\left(\left|\tilde{D}_{nk} - \frac{k}{n} \tilde{D}_{nn}\right| > \frac{C\sqrt{n}}{2}\right) + \mathbb{P}\left(\left|D_{nk} - \frac{k}{n} D_{nn} - [\tilde{D}_{nk} - \frac{k}{n} \tilde{D}_{nn}]\right| > \frac{C\sqrt{n}}{2}\right) \\ &\stackrel{n}{\ll} \frac{\mathbb{E}\left(\tilde{D}_{nk} - \frac{k}{n} \tilde{D}_{nn}\right)^2}{Cn} + \frac{n^{1-\theta}}{Cn}, \end{aligned}$$

which entails $\frac{1}{\sqrt{n}} |D_{nk} - \frac{k}{n} D_{nn}| = O_{\mathbb{P}}(1)$. W.l.o.g. assume $\Delta_n > 0$ for large n and observe that by (11) it holds $m_n(\tau+1) = \frac{\lfloor n\vartheta \rfloor}{\sqrt{n}} \frac{n - \lfloor n\vartheta \rfloor - 1}{n} \Delta_n \rightarrow +\infty$, as $n \rightarrow \infty$. Consequently, we have

$$\begin{aligned} \max_{k \leq n} \frac{1}{\sqrt{n}} \left|U_{nk} - \frac{k}{n} U_{nn}\right| &\geq \frac{1}{\sqrt{n}} \left|D_{n,\tau+1} - \frac{\tau+1}{n} D_{nn} + m_n(\tau+1)\right| \\ &\geq \frac{1}{\sqrt{n}} \left|D_{n,\tau+1} - \frac{\tau+1}{n} D_{nn} - |m_n(\tau+1)|\right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} +\infty. \end{aligned}$$

□

Proof of Theorem 3. Observe that $\mathbb{C}(Y_{ni}^{(\nu)}(\mathbf{b}), Y_{ni'}^{(\mu)}(\mathbf{c})) = \sum_{\ell=0}^{\infty} b_{n\ell}^{(\nu)} c_{n,i'-i+\ell}^{(\mu)} \sigma_{i-j}^2$ for $i \leq i'$ and arbitrary coefficient arrays \mathbf{b}, \mathbf{c} . By the strengthened decay condition, we have $\mathbb{E}(\mathbf{v}_n^\top \mathbf{Y}_{ni}(\mathbf{a}))^2 = O(1)$ uniformly in i , $\mathbb{E}\left(n^{-1/2} \sum_{i=1}^k \mathbf{v}_n^\top \mathbf{Y}_{ni}(\mathbf{a})\right)^2 = O(1)$ and

$$\mathbb{E}(\mathbf{v}_n^\top \bar{\mathbf{Y}}_n(\mathbf{a}))^2 = O\left(\|\mathbf{v}_n\|_{\ell_1}^2 \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{C}(Y_{ni}^{(\nu)}(\mathbf{a}), Y_{nj}^{(\nu)}(\mathbf{a}))\right) \stackrel{n}{\ll} n^{-1},$$

for $\mathbf{a} \in \{\mathbf{b}, \mathbf{c}\}$. It follows that

$$\mathbb{E}\left|\mathbf{w}_n^\top \bar{\mathbf{Y}}_n \sum_{i \leq k} \mathbf{v}_n^\top \mathbf{Y}_{ni}(\mathbf{a})\right| = \mathbb{E}\left|n^{-1/2} \sum_{i \leq k} \mathbf{v}_n^\top \mathbf{Y}_{ni} n^{-1/2} \sum_{i \leq k} \mathbf{w}_n^\top \mathbf{Y}_{ni}\right| = O(1)$$

and therefore

$$R_{ni}(\mathbf{a}) = -\mathbf{w}_n^\top \bar{\mathbf{Y}}_n(\mathbf{a}) \mathbf{v}_n^\top \mathbf{Y}_{ni}(\mathbf{a}) - \mathbf{v}_n^\top \bar{\mathbf{Y}}_n(\mathbf{a}) \mathbf{w}_n^\top \mathbf{Y}_{ni}(\mathbf{a}) + \mathbf{v}_n^\top \bar{\mathbf{Y}}_n(\mathbf{a}) \mathbf{w}_n^\top \bar{\mathbf{Y}}_n(\mathbf{a})$$

satisfies $\sup_{i \geq 1} \mathbb{E}|R_{ni}(\mathbf{a})| \ll n^{-1/2}$ and $\mathbb{E}\left|\sum_{i \leq k} R_{ni}(\mathbf{a})\right| = O(k/n)$. Put

$$\tilde{\xi}_{ni} = \tilde{\xi}_{ni}(\mathbf{v}_n^\top \mathbf{b}_n, \mathbf{v}_n^\top \mathbf{c}_n) = \begin{pmatrix} \tilde{Y}_{ni}(\mathbf{v}_n^\top \mathbf{b}_n) \tilde{Y}_{ni}(\mathbf{w}_n^\top \mathbf{b}_n) - \mathbb{E}[Y_{ni}(\mathbf{v}_n^\top \mathbf{b}_n) Y_{ni}(\mathbf{w}_n^\top \mathbf{b}_n)] \\ \tilde{Y}_{ni}(\mathbf{v}_n^\top \mathbf{c}_n) \tilde{Y}_{ni}(\mathbf{w}_n^\top \mathbf{c}_n) - \mathbb{E}[Y_{ni}(\mathbf{v}_n^\top \mathbf{c}_n) Y_{ni}(\mathbf{w}_n^\top \mathbf{c}_n)] \end{pmatrix},$$

where $\tilde{Y}_{ni}(\cdot) = Y_{ni}(\cdot) - \bar{Y}_n(\cdot)$. Since $\tilde{\xi}_{ni} = \xi_{ni} + \mathbf{R}_{ni}$, $\mathbf{R}_{ni} = (R_n(\mathbf{b}), R_n(\mathbf{c}))^\top$, we have for $k \leq n$ the estimates $\mathbb{E}\|\xi_{ni} - \tilde{\xi}_{ni}\|_\infty = \mathbb{E}\|\mathbf{R}_{ni}\|_\infty \ll n^{-1/2}$ uniformly in i . Consider the decomposition $\tilde{\mathbf{D}}_{nk} = \mathbf{D}_{nk} + \mathbf{R}_n$ if $\mathbf{R}_n = \sum_{i \leq k} \mathbf{R}_{ni}$. By Markov's inequality $\mathbb{P}(|\sum_{i \leq k} R_{ni}(\mathbf{a})| > \delta n^{\lambda'}) \leq \delta^{-1} n^{-\lambda'} \mathbb{E}|\sum_{i \leq k} R_{ni}(\mathbf{a})| = O(n^{-\lambda'})$, $\delta, \lambda' > 0$, such that $\mathbf{R}_n = o_{\mathbb{P}}(n^{\lambda'})$ for any $\lambda' > 0$. It follows that

$$\tilde{\mathbf{D}}_{nk} = \mathbf{D}_{nk} + \mathbf{R}_n, \quad \|\mathbf{R}_n\|_\infty = o_{\mathbb{P}}(n^{1/2}).$$

In view of (55) we may conclude that, on a new probability space for equivalent versions, $\|\mathbf{D}_{nt} - \mathbf{B}_n(t)\|_2 \leq C_n t^{1/2-\lambda}$, $t > 0$, a.s.. By virtue of [Billingsley, 1999, Sec. 21, Lemma 2] this strong approximation can be constructed on the original probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Consequently, we obtain $\|\tilde{\mathbf{D}}_{nt} - \mathbf{B}_n(t)\|_2 \leq C_n t^{1/2-\lambda} + o_{\mathbb{P}}(n^{1/2})$, $t > 0$, a.s.. Now it follows easily that assertions (i) and (ii) of Theorem 1 hold true with an additional error term $o_{\mathbb{P}}(n^{1/2})$ and (iii)-(vi) with an additional $o_{\mathbb{P}}(1)$ term. Finally, (vii) and (viii) hold in probability, if $C_n n^{-\lambda} = o(1)$. □

A.4. Proofs of Subsection 7.5.

Theorem 11. *Under the change-point alternative model (9) with $\tau = \lfloor n\vartheta \rfloor$, $\vartheta \in (0, 1)$, there exist a \mathcal{F}_{nk} -martingale array \tilde{D}_{nk} , $1 \leq k, n \geq 1$, such that*

$$(82) \quad \mathbb{E}(D_{nk} - \tilde{D}_{nk})^2 \stackrel{n,k}{\ll} k^{1-\theta}.$$

and hence for $k \leq n$ and $n \geq 1$

$$(83) \quad \mathbb{E} \left(D_{nk} - \frac{k}{n} D_{nn} - [\tilde{D}_{nk} - \frac{k}{n} \tilde{D}_{nn}] \right)^2 \stackrel{n,k}{\ll} k^{1-\theta}.$$

Further, if $0 < \beta < 1/2$, then

$$(84) \quad \mathbb{E} \left(\left(\frac{n}{k} \right)^\beta \left| D_{nk} - \frac{k}{n} D_{nn} - [\tilde{D}_{nk} - \frac{k}{n} \tilde{D}_{nn}] \right| \right)^2 \stackrel{n,k}{\ll} n^{1-\theta}.$$

Proof of Theorem 11. Recall (72) and put for each $n \geq 1$

$$(85) \quad \tilde{D}_{nk} = \delta M_k^{(n)}(\mathbf{b}) \mathbf{1}(k \leq \tau) + [\delta M_\tau^{(n)}(\mathbf{b}) + \delta M_k^{(n)}(\mathbf{c}) - \delta M_\tau^{(n)}(\mathbf{c})] \mathbf{1}(k > \tau), \quad k \geq 1.$$

It is clear that $\mathbb{E}(\tilde{D}_{nk} | \mathcal{F}_{n,k-1}) = 0$ holds if $k \leq \tau$ and $k > \tau + 1$. In addition, for $k = \tau + 1$ we have

$$\begin{aligned} \mathbb{E}[\tilde{D}_{nk} - \tilde{D}_{n,k-1} | \mathcal{F}_{n,k-1}] &= \mathbb{E}[\delta M_\tau^{(n)}(\mathbf{b}) + \delta M_{\tau+1}^{(n)}(\mathbf{c}) - \delta M_\tau^{(n)}(\mathbf{c}) - \delta M_\tau^{(n)}(\mathbf{b}) | \mathcal{F}_{n,\tau}] \\ &= \mathbb{E}[M_{n,\tau+1}^{(n)}(\mathbf{c}) - M_{n,\tau}^{(n)}(\mathbf{c}) | \mathcal{F}_{n,\tau}] = 0, \end{aligned}$$

because $M_{nk}^{(n)}(\mathbf{c})$ is a \mathcal{F}_{nk} -martingale array. Since

$$D_{nk} = \mathbf{D}_{nk}^{(1)} \mathbf{1}(k \leq \tau) + [\mathbf{D}_{n\tau}^{(1)} + \mathbf{D}_{nk}^{(2)} - \mathbf{D}_{n\tau}^{(2)}] \mathbf{1}(k > \tau),$$

the triangle inequality provides the upper bound

$$\|\mathbf{D}_{nk}^{(1)} - \delta M_k^{(n)}(\mathbf{b})\|_{L_2} + \|\mathbf{D}_{n\tau}^{(1)} - \delta M_\tau^{(n)}(\mathbf{b})\|_{L_2} + \|\mathbf{D}_{nk}^{(2)} - \delta M_k^{(n)}(\mathbf{c})\|_{L_2} + \|\mathbf{D}_{n\tau}^{(2)} - \delta M_\tau^{(n)}(\mathbf{c})\|_{L_2}$$

for $\|D_{nk} - \tilde{D}_{nk}\|_{L_2}$, which is $\ll k^{1/2-\theta/2}$ by virtue of Lemma 2, see (49) and (50). This verifies (82), i.e.,

$$\|D_{nk} - \tilde{D}_{nk}\|_{L_2} \stackrel{n,k}{\ll} k^{1/2-\theta/2}.$$

As a consequence, for $k \leq n$ and $n \geq 1$

$$\left\| D_{nk} - \frac{k}{n} D_{nn} - [\tilde{D}_{nk} - \frac{k}{n} \tilde{D}_{nn}] \right\|_{L_2} \stackrel{k,n}{\ll} \|D_{nk} - \tilde{D}_{nk}\|_{L_2} + \frac{k}{n} \|D_{nn} - \tilde{D}_{nn}\|_{L_2} \stackrel{k,n}{\ll} k^{1/2-\theta/2}.$$

Lastly, since we may assume that θ is small enough to ensure $\theta + 2\beta < 1$, it holds for $1 \leq k \leq n$

$$(86) \quad \mathbb{E} \left(\left(\frac{n}{k} \right)^\beta \left| D_{nk} - \frac{k}{n} D_{nn} - [\tilde{D}_{nk} - \frac{k}{n} \tilde{D}_{nn}] \right| \right)^2 \stackrel{n,k}{\ll} n^{2\beta} k^{-2\beta} k^{1-\theta} \leq n^{1-\theta},$$

□

Proof of Theorem 7. The proof is completed by considering the decomposition

$$\begin{aligned} \sum_{|h| \leq m_n} w_{mh} \tilde{\Gamma}_n(u; h, d) - \alpha^2(u, \mathbf{b}, \mathbf{c}) &= \sum_{|h| \leq m_n} w_{mh} \tilde{\Gamma}_n(u; h, d) - \sum_{h \in \mathbb{Z}} \Gamma(u; h, d) + o(1) \\ &= A_n(u; d) + B_n(u; d) + C_n(u; d) + D_n(u; d) + o(1), \end{aligned}$$

where the $o(1)$ term is uniform over $d \in \mathbb{N}$ and

$$\begin{aligned} A_n(u; d) &= \sum_{|h| \leq m_n} w_{mh} [\tilde{\Gamma}_n(u; h, d) - \mathbb{E}(\tilde{\Gamma}_n(u; h, d))], \quad B_n(u; d) = \sum_{|h| \leq m_n} w_{mh} [\mathbb{E}(\tilde{\Gamma}_n(u; h, d)) - \Gamma(u; h, d)], \\ C_n(u; d) &= \sum_{|h| \leq m_n} [w_{mh} - 1] \Gamma(u; h, d), \quad D_n(u; d) = - \sum_{|h| > m_n} \Gamma(u; h, d). \end{aligned}$$

$B_n(u; d)$ has been already estimated in (60) and (62) implies $\sup_{u \in [\varepsilon, 1]} \sup_{d \in \mathbb{N}} |D_n(u; h, d)| = o(1)$, as $n \rightarrow \infty$. Denote the counting measure on \mathbb{Z} by $d\nu$. Then by Fubini and (65)

$$\begin{aligned} \sup_{u \in [\varepsilon, 1]} \sup_{d \in \mathbb{N}} \mathbb{E}|A_n(u; d)| &\leq \sup_{u \in [\varepsilon, 1]} \sup_{d \in \mathbb{N}} \int_{\mathbb{Z}} w_{mh} \mathbb{E}|\tilde{\Gamma}_n(u; h, d) - \mathbb{E}\tilde{\Gamma}_n(u; h, d)| \mathbf{1}(|h| \leq m_n) d\nu(h) \\ &\leq 2Wm_n \sup_{u \in [\varepsilon, 1]} \sup_{d \in \mathbb{N}} \max_{|h| \leq m_n} \mathbb{E}|\tilde{\Gamma}_n(u; h, d) - \mathbb{E}\tilde{\Gamma}_n(u; h, d)| = o(1). \end{aligned}$$

Lastly, $\sup_{u \in [\varepsilon, 1]} \sup_{d \in \mathbb{N}} |C_n(u; h, d)| = o(1)$, as $n \rightarrow \infty$, follows by dominated convergence. \square

ACKNOWLEDGMENTS

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

REFERENCES

- [Aue et al., 2009] Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, 37(6B):4046–4087.
- [Avanesov and Buzun, 2018] Avanesov, V. and Buzun, N. (2018). Change-point detection in high-dimensional covariance structure. *arxiv:1610.03783*.
- [Bar-Joseph et al., 2012] Bar-Joseph, Z., Gitter, A., and Simon, I. (2012). Studying and modelling dynamic biological processes using time-series gene expression data. *Nature Reviews Genetics*, 13:552–564.
- [Berkas et al., 2009] Berkas, I., Gombay, E., and Horváth, L. (2009). Testing for changes in the covariance structure of linear processes. *J. Statist. Plann. Inference*, 139(6):2044–2063.
- [Billingsley, 1999] Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition. A Wiley-Interscience Publication.
- [Birnbaum et al., 2013] Birnbaum, A., Johnstone, I. M., Nadler, B., and Paul, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.*, 41(3):1055–1084.
- [Breitung and Eickmeier, 2011] Breitung, J. and Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *J. Econometrics*, 163(1):71–84.
- [Brodie et al., 2009] Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009). Sparse and stable Markowitz portfolios. *Proceedings the National Academy of Sciences of the United States of America*, 106(30):12267–12272.
- [Cai et al., 2015] Cai, T., Ma, Z., and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields*, 161(3-4):781–815.
- [Cho and Fryzlewicz, 2015] Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 77(2):475–507.
- [Csörgő and Horváth, 1993] Csörgő, M. and Horváth, L. (1993). *Weighted approximations in probability and statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester. With a foreword by David Kendall.
- [Csörgő and Horváth, 1997] Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. With a foreword by David Kendall.
- [Davis and Kahan, 1970] Davis, C. and Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46.
- [Erichson et al., 2018] Erichson, N. B., Zheng, P., Manohar, K., Brunton, S. L., Kutz, J. N., and Aravkin, A. A. (2018). Sparse principal component analysis via variable projection. *arXiv:1804.00341*.
- [Ferber, 2018] Ferber, D. (2018). On the supremum of a Brownian bridge standardized by its maximizing point with applications to statistics. *Statist. Probab. Lett.*, 134:63–69.
- [Galeano and Peña, 2007] Galeano, P. and Peña, D. (2007). Covariance changes detection in multivariate time series. *J. Statist. Plann. Inference*, 137(1):194–211.
- [Han and Inoue, 2015] Han, X. and Inoue, A. (2015). Tests for parameter instability in dynamic factor models. *Economet. Theor.*, 31(5):1117–1152.
- [Horváth and Rice, 2019] Horváth, L. and Rice, G. (2019). Asymptotics for empirical eigenvalue processes in high-dimensional linear factor models. *J. Multivariate Anal.*, 169:138–165.
- [Johnson and Lindenstrauss, 1984] Johnson, W. B. and Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI.
- [Johnstone and Lu, 2009] Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.*, 104(486):682–693.
- [Jolliffe et al., 2003] Jolliffe, L., Trendafilov, N., and Uddin, M. (2003). A modified principal component technique based on the lasso. *J. Comput. Graph. Statist.*, 12:531–547.

- [Jung and Marron, 2009] Jung, S. and Marron, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.*, 37(6B):4104–4130.
- [Kane and Nelson, 2014] Kane, D. M. and Nelson, J. (2014). Sparser Johnson-Lindenstrauss transforms. *J. ACM*, 61(1):Art. 4, 23.
- [Kouritzin, 1995] Kouritzin, M. A. (1995). Strong approximation for cross-covariances of linear variables with long-range dependence. *Stochastic Process. Appl.*, 60(2):343–353.
- [Ledoit and Wolf, 2004] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.*, 88(2):365–411.
- [Li and Zhao, 2013] Li, X. and Zhao, Z. (2013). Testing for changes in autocovariances of nonparametric time series models. *J. Statist. Plann. Inference*, 143(2):237–250.
- [Ma, 2013] Ma, Z. (2013). Sparse principal component analysis and iterative thresholding. *Ann. Statist.*, 41(2):772–801.
- [Na et al., 2011] Na, O., Lee, Y., and Lee, S. (2011). Monitoring parameter change in time series models. *Stat. Methods Appl.*, 20(2):171–199.
- [Paul, 2007] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica*, 17(4):1617–1642.
- [Paul and Johnstone, 2007] Paul, D. and Johnstone, I. M. (2007). Augmented sparse principal component analysis for high dimensional data. Technical Report arXiv:1202.1242.
- [Philipp, 1986] Philipp, W. (1986). A note on the almost sure approximation of weakly dependent random variables. *Monatsh. Math.*, 102(3):227–236.
- [Sancetta, 2008] Sancetta, A. (2008). Sample covariance shrinkage for high dimensional dependent data. *J. Multivariate Anal.*, 99(5):949–967.
- [Schweinberger et al., 2017] Schweinberger, M., Babkin, S., and Ensor, K. B. (2017). High-dimensional multivariate time series with additional structure. *J. Comput. Graph. Statist.*, 26(3):610–622.
- [Shen et al., 2013] Shen, D., Shen, H., and Marron, J. S. (2013). Consistency of sparse PCA in high dimension, low sample size contexts. *J. Multivariate Anal.*, 115:317–333.
- [Shorack and Wellner, 1986] Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
- [Steland, 2005] Steland, A. (2005). Random walks with drift—a sequential approach. *J. Time Ser. Anal.*, 26(6):917–942.
- [Steland, 2018] Steland, A. (2018). Shrinkage for covariance estimation: Asymptotics, confidence intervals, bounds and applications in sensor monitoring and finance. *Statistical Papers*, 59:1441–1462.
- [Steland and von Sachs, 2017] Steland, A. and von Sachs, R. (2017). Large-sample approximations for variance-covariance matrices of high-dimensional time series. *Bernoulli*, 23(4A):2299–2329.
- [Steland and von Sachs, 2018] Steland, A. and von Sachs, R. (2018). Asymptotics for high-dimensional covariance matrices and quadratic forms with applications to the trace functional and shrinkage. *Stochastic Process. Appl.*, 128(8):2816–2855.
- [van der Vaart, 1998] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- [Vostrikova, 1981] Vostrikova, L. J. (1981). Discovery of disorder in multidimensional random processes. *Dokl. Akad. Nauk SSSR*, 259(2):270–274.
- [Wang and Samworth, 2018] Wang, T. and Samworth, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 80(1):57–83.
- [Wang and Fan, 2017] Wang, W. and Fan, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.*, 45(3):1342–1374.
- [Yu et al., 2015] Yu, Y., Wang, T., and Samworth, R. J. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323.