Received: 21 November 2019 | Revised: 8 October 2020 | Accepted: 11 November 2020

DOI: 10.1111/sjos.12508

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Large-sample approximations and change testing for high-dimensional covariance matrices of multivariate linear time series and factor models

Monika Bours | **Ansgar Steland**

Institute of Statistics, RWTH Aachen University, Aachen, Germany

Correspondence
Ansgar Steland, Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany.
Email: steland@stochastik.rwth-aachen.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: STE 1134-11/1, 1134-11/2

Abstract

Various statistical problems can be formulated in terms of a bilinear form of the covariance matrix. Examples are testing whether coordinates of a high-dimensional random vector are uncorrelated, constructing confidence intervals for the risk of optimal portfolios or testing for the stability of a covariance matrix, especially for factor models. Extending previous works to a general high-dimensional multivariate linear process framework and factor models, we establish distributional approximations for the associated bilinear form of the sample covariance matrix. These approximations hold for increasing dimension without any constraint relative to the sample size. The results are used to construct change-point tests for the covariance structure, especially in order to check the stability of a high-dimensional factor model. Tests based on the cumulated sum (CUSUM), self-standardized CUSUM and the CUSUM statistic maximized over all subsamples are considered. Size and power of the proposed testing methodology are investigated by a simulation study and

[Correction added on 22 February 2021, after first online publication: Funding Statement has been added.]

illustrated by analyzing the Fama and French factors for a change due to the SARS-CoV-2 pandemic.

KEYWORDS

change-points, COVID-19 crisis, factor model, high-dimensional statistics, multivariate analysis, portfolio analysis, strong approximation, time series

1 | INTRODUCTION

In this article, we study large sample approximations of bilinear forms of the sample covariance matrix of a d-dimensional time series of length n within a high-dimensional framework, which can serve as a basis for various inferential procedures such as the analysis of dependencies of observed variables in terms of statistical tests and confidence intervals or change detection. Areas where this setting is of importance are diverse and comprise genetics, environmental statistics as well as econometrics and finance. To give a concrete example, in finance optimal portfolio selection is a classical problem directly pointing to a quadratic form. Here, for a random vector $\mathbf{Y}$ of returns with covariance matrix $\boldsymbol{\Sigma}$, one aims at determining a variance-minimizing portfolio $\boldsymbol{w}^*$ under additional constraints (such as no short sales) leading to the risk $\operatorname{Var}(\mathbf{w}^*) = \mathbf{w}^{*\prime}\boldsymbol{\Sigma}\mathbf{w}^*$ of the optimal portfolio $\mathbf{w}^*$. Imposing sparsity constraints such as $\|\mathbf{w}\|_{\ell_q}$, $q \in \{0, 1\}$, see Brodie et al. (2009), may lead to substantially reduced transaction costs and allows for cheap index tracking if the assets belong to an index. The associated minimized financial risk can be estimated by the bilinear form $\mathbf{w}^{*\prime}\hat{\boldsymbol{\Sigma}}_n\mathbf{w}^*$ using the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_n$ from a d-dimensional time series of length n. Clearly, it is of interest to make inference on that quadratic form and to detect change-points where the covariance matrix and hence the risk changes. Another interesting econometric problem is to detect changes (breaks) in economical or statistical factor models, which can be done by investigating the eigenstructure or related expansions. We shall illustrate this by analyzing the well-known Fama and French factors for a change due to the COVID-19 coronavirus pandemic.

The proposed testing methodology allows for high-dimensional data intractable by other known methods such as Aue et al. (2009), Han and Inoue (2015), or Kao et al. (2018), and is extremely cheap in terms of computational costs. On the other hand, whereas other tests are based on (quadratic forms of) cumulated sums for $\operatorname{vech}(\mathbf{Y}_t\mathbf{Y}_t')$ or linear mappings of those vectors, our approach is to study bilinear forms of matrix-valued CUSUMs corresponding to projections $\mathbf{v}'\mathbf{Y}_t$ of the data. By construction and purpose, the methods allow for large d including $d \geq n$ and do not require to invert a matrix of dimension $d \times d$ as many other approaches, which can become intractable in applications even when $d < n$.

We contribute Gaussian approximations for bilinear forms associated to the sample covariance matrix under general multivariate linear time series including a rich class of factor models. These results complement previous works of Kouritzin (1995), Steland and von Sachs (2017, 2018), and Steland (2020). As in these papers, the Gaussian approximations hold true without any constraint on the dimension d, which may grow arbitrarily with the sample size n, and allow to construct various change-point tests. On the other hand Steland (2020) studies unweighted as well as weighted CUSUM test statistics for a high-dimensional linear time series framework covering multivariate linear processes with approximately white noise innovations and focuses on VARMA and spiked covariance models, here we allow for nonapproximate multivariate linear time series and consider the unweighted CUSUM, a self-standardized CUSUM as well as a maximized CUSUM statistic, where maximization is over all available subsamples. The latter two statistics have certain merits and were not studied in the above cited papers. Self-normalization is an approach to avoid estimation of long-run-variance parameters, but it changes the asymptotic law which impedes the calculation of critical values. Instead of a consistent estimator for the nuisance parameter one uses a weakly convergent statistic, whose limiting law is proportional to the parameter. Hence, after self-normalization the nuisance parameter cancels in the limit. This approach has been proposed by Kiefer et al. (2000) for dynamic regression inference, but the phenomenon and its merits are also well known in unit root testing and monitoring, see Kwiatkowski et al. (1992), Breitung (2002), and Steland (2007), among others. Here we basically use the self-normalizing statistic of Shao and Zhang (2010) which takes into account the presence of a change-point. The CUSUM statistic maximized all over subsamples addresses the fact that the classical (unweighted) CUSUM test targets the at most one change-point alternative hypothesis. By considering all possible subsamples and taking the maximum of all corresponding CUSUMs allows to detect complex change-point alternatives, which may be undetectable by the classical unweighted CUSUM statistic (see the simulations for an example).

This article focuses to some extent on factor models, which have not been treated in the above cited papers. Factor models are regarded as a powerful tool for modeling and understanding the common dependence among multivariate outputs. They are widely used in various areas including econometrics and finance to model macroeconomic variables such as the GDP and inflation, see Bai and Ng (2008), or to model asset and portfolio returns from financial markets, see Goldfarb and Iyengar (2003), Johnstone (2001), and the classical Fama approach (Fama & French, 1993, 2015). In medicine they are used to explain genetic data, in environmetrics to model complex spatial–temporal dependencies. Their use in natural sciences and technology is also widespread, see, for example, Reyment and Jvreskog (1996). The class of factor models studied here covers the case of an arbitrary but finite number of independent factors and the case of a growing or even infinite number of factors, which are generally correlated but may be independent as well (under certain constraints). We especially elaborate on testing the stability of a factor model as a relevant special case in many areas. This problem is approached by considering certain pseudo-spectral representations and to test for a change with respect to the (leading) pseudo-eigenvectors by applying a CUSUM change-point test to a quadratic form calculated using pseudo-eigenvectors. More detailed analyses using a large number of pseudo-eigenvectors can be easily conducted by a multiple testing procedure. This method, described in detail in Section 2, shows decent power in simulations despite its simplicity and will be illustrated by a real data analysis. Especially, self-standardization often improves the accuracy of the type I error rate, although the unweighted CUSUM with estimated variance tends to be more powerful for larger deviations from the no-change null hypothesis. Self-standardization also benefits in terms of detection power when the number of projections used for testing is increased. The CUSUM maximized over subsamples leads to a loss of power for the classical at most one change-point model, but can be very powerful for complex change scenarios where other tests completely fail.

The proposed class of change-point tests considering bilinear forms of the sample covariance matrix requires to select projection (weighting) vectors. For several problems their choice is obvious. This applies, for example, to testing candidate coordinates against remaining ones (as especially arising in statistical genetics), or when considering sparse optimal portfolio selection mentioned already above, where ℓ_1-bounded portfolio vectors are constructed from historical data sets. A quite general approach is to use (estimated) eigenvectors of the covariance matrix, in order to test the stability of the covariance structure. Calculating the eigenvalue–eigenvector pairs from the sample covariance matrix is equivalent to determining the spectral decomposition, which represents the sample covariance matrix as a linear combination of outer products of the eigenvectors with weights given by the eigenvalues. Somewhat more general, any representation of an estimator of the covariance matrix as a linear combination of outer products can be interpreted as an estimation of the (pseudo) eigenvector–eigenvalue pairs—and used for our testing approach. Testing the stability of the covariance structure in this way also offers an attractive approach to test for the stability of a factor model structure, and we shall elaborate this idea in some detail.

As in practice the above testing methodology usually relies on an estimator of the covariance matrix, it is worth discussing a few approaches from the rich literature on this subject. The main issue is the well-known fact that the sample covariance matrix behaves poor when d is large. For $d > n$ it is singular and hence not invertible, and even when $d < n$ but d/n is close to 1, the estimated eigenstructure tends to be systematically distorted, as the small eigenvalues are underestimated and large eigenvalues overestimated. This results in a numerically ill-conditioned estimator, which means that inversion amplifies the estimation error dramatically. Classical results, addressing Gaussian samples, to overcome this issue are Dempster (1972), where selected entries are set to zero, and Stein (1956, 1975), who shrinks the eigenvalues toward a common value. Shrinkage methods have been later thoroughly investigated by Ledoit and Wolf (2003, 2004) for i.i.d. samples and by Sancetta (2008) for the weak-dependent case. Second, one could use Bayesian and empirical Bayes estimators, which are related to the shrinkage estimators but provide alternative interpretations. For example, Chen (1979) and Daniels and Kass (1999, 2001), discuss priors such that the correlations are shrunken to 0 and the covariance matrix toward the identity, Yang and Berger (1994) develop the reference prior approach, and Barnard et al. (2000) put a prior on the standard deviations and model the correlations given standard deviations by an inverse-Wishart distribution. Third, one could use the method based on random matrix theory to denoise the sample covariance matrix, as studied by Laloux et al. (2000), Sharifi et al. (2004), and Pafka and Kondor (2003). For a selective overview on high-dimensional factor model-based covariance estimation for both observable and latent factors we refer to Fan et al. (2013, 2016), Bai and Shi (2011), and the references therein. Inferential methods for factor models in large dimensions have been studied by Bai (2003). He considered the principal components estimator and derived the rate of convergence and the limiting distributions of the estimated factors and factor loadings. Bai and Li (2012, 2016) studied maximum likelihood estimation. Their results cover consistency, the rate of convergence, and the limiting distributions for different sets of identification conditions. When the factors are observable, factor models indeed allow for more efficient estimators of the covariance matrix, see Fan et al. (2008, 2011).

The specific problem to test for a change in factor models has been discussed by several authors, for example, Breitung and Eickmeier (2011), Han and Inoue (2015) and, recently, by Kao et al. (2018), among others. Breitung and Eickmeier (2011) consider assumptions as in Bai (2003), especially a finite number of factors and asymptotics for panel data with $n/d^2 = o(1)$ where d stands for the size of the panel. They study tests based on the principal component estimators of the factor model. Kao et al. (2018) consider a specialized CUSUM change-point test for a d-dimensional time series based on estimates of the eigenvalue and eigenvector pairs and develop asymptotic theory for factor models with innovations following a nonlinear specification going beyond the setting studied here. But their approach is restricted to the low-dimensional framework of a fixed dimension d. Furthermore, their test statistic needs the condition $d/n < 1/2$ to be computable and requires to invert a $d \times d$ matrix for standardization. Although standardization typically improves statistical power, it may be infeasible when analyzing high-dimensional big data. The approach of Han and Inoue (2015) even requires to calculate a $d(d+1)/2 \times d(d+1)/2$ dimensional matrix to conduct the test. Contrary, the asymptotics studied here allows for (essentially unconstrained) increasing dimension as well as for an increasing number of factors (including an infinite number of factors), and the proposed CUSUM-based procedures are highly efficient from a computational point of view. Furthermore, the method is more general and not restricted to analyses of the eigenstructure of a factor model. For example, by simply choosing appropriate weighting vectors, one may easily detect changes in covariances of (sets of) coordinates as well. As demonstrated in the simulations, the tests can compete with other approaches in terms of accuracy and detection power.

The organization of the article is as follows. Section 2 explains the general setting, discusses its basic relationship to projection-based analyses, and introduces the bilinear form of interest. Notation, the specific model for the vector time series and its interpretation in terms of an infinite-dimensional factor model with unobservable factors as well as assumptions are introduced and discussed in Section 3. Section 4 provides several results on strong approximations justifying the proposed inferential procedures. We also propose estimators for the asymptotic variance parameters and show their consistency, uniformly in the dimension. The finite sample properties in terms of size and power of the tests are investigated to some extent in Section 5 by a simulation study. We also illustrate the method by analyzing Fama and French factors and find that there is evidence for a change due to the COVID-19 coronavirus pandemic in the whole eigenstructure which can be dated at the end of February 2020. Proofs of main results are provided in Section 6.

2 | MODEL, ASSUMPTIONS AND EXAMPLES

The general approximation result for partial sums holds for sparse projection vectors $\mathbf{v}_n, \mathbf{w}_n$ in the sense that they have uniformly bounded ℓ_1-norm. For appropriately standardized partial sums, however, one can employ weighting vectors with uniformly bounded ℓ_2-norms.

Assumption 1. (i) Let $\mathbf{w}_n = (w_1, \dots, w_{d_n})'$, $n \geq 1$, be a sequence of weights $w_j = w_{nj}$, not necessarily nonnegative, with uniformly bounded ℓ_1-norm, that is,

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

(ii) $\mathbf{w}_n = (w_1, \dots, w_{d_n})'$, $n \geq 1$, is a sequence of weights $w_j = w_{nj}$, with

$$\sup_{n\in\mathbb{N}} \|\mathbf{w}_n\|_{\ell_2} = \sup_{n\in\mathbb{N}} \sqrt{\sum_{\nu=1}^{d_n} w_\nu^2} < \infty. \tag{2}$$

It is worth briefly discussing the first assumption. Sequences $\{w_i : i \in \mathbb{N}\}$ with $\sum_{i=1}^{\infty} |w_i| < \infty$ certainly satisfy Assumption 1 (i). Averaging a finite number of coordinates is also covered. Furthermore, this assumption allows for weights that depend on the dimension d_n. For example, one may average all coordinates by using the weights $w_{ni} = \frac{1}{d_n}$, for $i = 1, \dots, d_n$. Although the dependence on the sample size n through the dimension d_n may be of primary importance for high-dimensional problems, several of our results even allow the weights to depend on n.

Let us assume the following linear process framework. The νth coordinate of the observed d_n-dimensional random vector $\mathbf{Y}_{ni}$, $1 \le i \le n$, is given by a sum (equivalently a linear combination) of $L \in \mathbb{N}$ linear processes

$$Y_i^{(\nu)} = \sum_{l=1}^{L} \sum_{j=0}^{\infty} c_{nj}^{(\nu,l)} \varepsilon_{i-j}^{(l)}, \tag{3}$$

where $\{c_{nj}^{(\nu,l)} : j \ge 0\}$ are coefficients (possibly depending on n) and $\{\varepsilon_i^{(l)} : i \in \mathbb{Z}\}$, $l = 1, \dots, L$, are mean zero error terms. Observe that model (3) can be reformulated as a multivariate linear process

$$\mathbf{Y}_{ni} = \sum_{j=0}^{\infty} \mathbf{C}_{nj} \boldsymbol{\varepsilon}_{i-j},$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_i^{(1)}, \dots, \varepsilon_i^{(L)})'$, $i \ge 1$, are i.i.d. random vectors with zero mean and the $d_n \times L$ coefficient matrices are given by $\mathbf{C}_{nj} = (\mathbf{c}_{nj}^{(1)}, \dots, \mathbf{c}_{nj}^{(d_n)})'$ with $\mathbf{c}_{nj}^{(\nu)} = (c_{nj}^{(\nu,1)}, \dots, c_{nj}^{(\nu,L)})'$.

Assumption 2. The $L \in \mathbb{N}$ innovation processes $\{\varepsilon_i^{(1)} : i \in \mathbb{Z}\}, \dots, \{\varepsilon_i^{(L)} : i \in \mathbb{Z}\}$ are stochastically independent.

Throughout the article we assume the uniform moment condition

$$\sup_i \mathrm{E}[|\varepsilon_i^{(l)}|^{4+\delta}] < \infty, \quad l = 1, \dots, L, \text{for some } \delta > 0. \tag{4}$$

The existence of higher moments is assumed where needed. Similar to Steland and von Sachs (2017) we impose the following decay condition on the coefficients.

Assumption 3. $\{c_{nj}^{(\nu,l)} : j \in \mathbb{N}_0\}$, $\nu = 1, \dots, d$, $l = 1, \dots, L$, satisfy

$$\sup_{n \in \mathbb{N}} \max_{1 \le \nu \le d_n} |c_{nj}^{(\nu,l)}|^2 \ll (j \vee 1)^{-\frac{3}{2}-\vartheta}, \quad \forall\, l = 1, \dots, L,$$

for some $0 < \vartheta < \frac{1}{2}$.

Here and in the sequel $a_n \ll b_n$ stands for $a_n = O(b_n)$. Furthermore, we shall write $a_{nm} \overset{n,m}{\ll} b_{nm}$ if there exists a constant C such that $a_{nm} \le C b_{nm}$ for all n, m.

The following result shows that a large class of general multivariate linear processes with correlated coordinates of the innovations are a special case of our model, as they can be represented as $\boldsymbol{\Sigma}_L^{1/2} \boldsymbol{\varepsilon}_i$ for some covariance matrix $\boldsymbol{\Sigma}_L$. The following result shows that a weak sparsity assumption on the left-singular vectors of $\boldsymbol{S}_L = \boldsymbol{\Sigma}_L^{1/2}$ is sufficient.

Lemma 1.

(i) Let $\mathbf{G}$ *be a* $d_n \times L$ *matrix with uniformly* ℓ_1*-bounded left-singular vectors. Then the class of multivariate linear processes of the form*

$$\mathbf{Y}_{ni} = \sum_{j=0}^{\infty} (\mathbf{C}_{nj} | \mathbf{0}_{d_n \times (d_n - L)}) \mathbf{G} \boldsymbol{\varepsilon}_{i-j}, \quad i \ge 0,$$

for L-dimensional i.i.d. innovations $\{\boldsymbol{\varepsilon}_i : i \ge 1\}$ *with independent coordinates satisfies Assumption 3.*

(ii) The class of multivariate linear processes of the form

$$\mathbf{Y}_{ni} = \sum_{j=0}^{\infty} \mathbf{C}_{nj}\mathbf{S}_L\boldsymbol{\epsilon}_{i-j}, \quad i \geq 0,$$

for some regular matrix $\mathbf{S}_L$ with uniformly ℓ_1-bounded left-singular vectors and L-dimensional i.i.d. innovations $\{\boldsymbol{\epsilon}_i : i \geq 1\}$ with independent coordinates, satisfies Assumption 3.

Example 1 (Factor models with a finite number of independent factors). Recall that a factor model for a d-dimensional random vector $\mathbf{Y}$ assumes that

$$\mathbf{Y} = \mathbf{B}\mathbf{F} + \mathbf{E}$$

with

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_1^{(1)} & \dots & b_1^{(K)} \\ \vdots & & \\ b_d^{(1)} & \dots & b_d^{(K)} \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} F_1 \\ \vdots \\ F_K \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} E_1 \\ \vdots \\ E_d \end{pmatrix},$$

where the random variables $F_1, \dots, F_K$ represent K uncorrelated factors, $\mathbf{B}$ is the matrix of the factor loadings $b_i^{(j)}$, and $E_1, \dots, E_d$ are d idiosyncratic errors, which are assumed to be independent from $F_1, \dots, F_K$. Classical factor analysis assumes a fixed dimension d, while the sample size n is allowed to increase. The main results of this article allow for factor models of large dimensions, where the dimension $d = d_n$ grows with the sample size n and can even grow faster than n. Furthermore, the coefficients of the linear series may depend on n, which we therefore now make explicit. The covariance matrix of $\mathbf{Y}_n$ is given by

$$\boldsymbol{\Sigma}_n = \mathbf{B}_n \mathrm{Cov}(\mathbf{F}_n)\mathbf{B}_n' + \mathrm{Cov}(\mathbf{E}_n).$$

A factor models assumes that a more or less large part of the dependence structure can be explained by the factors and reduces the dimensionality of the covariance matrix in this way. Given the factors, the remaining correlation between the coordinates should be small, suggesting that a sparsity assumption for the covariance matrix of the errors $\mathbf{E}_n$ is reasonable. Here, one often even assumes that $\mathrm{Cov}(\mathbf{E}_n)$ is a diagonal matrix, otherwise a so-called approximate factor model is present.

Now suppose that the above factor model is observed at n time instants yielding the vector time series $\mathbf{Y}_{ni} = (Y_{ni}^{(1)}, \dots, Y_{ni}^{(d_n)})'$, $1 \leq i \leq n$,

$$\mathbf{Y}_{ni} = \mathbf{B}_n \mathbf{F}_{ni} + \mathbf{E}_{ni},$$

for $i = 1, \dots, n$, where the vector $\mathbf{F}_{ni} = (F_{ni}^{(1)}, \dots, F_{ni}^{(K)})'$ contains the values of all K factors at time i, for $i = 1, \dots, n$, and the matrix $\mathbf{B}_n$ is as above. The vector $\mathbf{E}_{ni} = (E_{ni}^{(1)}, \dots, E_{ni}^{(d_n)})'$ is an idiosyncratic error component that contains the noise due to the d_n variables at time i and not explained by the factors, for $i = 1, \dots, n$.

Specifically, with $L=K+1$ we assume that the coordinates of the factors as well as the idiosyncratic errors are linear processes given by

$$E_{ni}^{(\nu)} = \sum_{j=0}^{\infty} c_{nj}^{(\nu)} \epsilon_{i-j}^{(L)} \quad \text{and} \quad F_{ni}^{(l)} = \sum_{j=0}^{\infty} g_{nj}^{(l)} \epsilon_{i-j}^{(l)}, \quad \ell = 1, \dots, K, \tag{5}$$

for coefficients $\{c_{nj}^{(\nu)} : j \geq 0\}$ and $\{g_{nj}^{(l)} : j \geq 0\}$, for $\nu = 1, \dots, d_n$, $l = 1, \dots, K$. Then the νth coordinate of $\mathbf{Y}_{ni}$, $1 \leq i \leq n$, is given by

$$Y_{ni}^{(\nu)} = \mathbf{b}_{n\nu}' \mathbf{F}_{ni} + E_{ni}^{(\nu)} = \sum_{l=1}^{L-1} b_{\nu}^{(l)} F_{ni}^{(l)} + E_{ni}^{(\nu)}, \quad \text{for } i = 1, \dots, n, \tag{6}$$

where $\mathbf{b}_{n\nu}'$ denotes the νth row of $\mathbf{B}_n$. If we define $c_{nj}^{(\nu,l)} = b_{\nu}^{(l)} g_{nj}^{(l)}$, for $l=1, \dots L-1$, and $c_{nj}^{(\nu)} = c_{nj}^{(\nu,L)}$, then it follows

$$\begin{aligned} Y_{ni}^{(\nu)} &= \sum_{j=0}^{\infty} c_{nj}^{(\nu,1)} \epsilon_{i-j}^{(1)} + \cdots + \sum_{j=0}^{\infty} c_{nj}^{(\nu,L-1)} \epsilon_{i-j}^{(L-1)} + \sum_{j=0}^{\infty} c_{nj}^{(\nu,L)} \epsilon_{i-j}^{(L)} \\ &= \sum_{l=1}^{L} \sum_{j=0}^{\infty} c_{nj}^{(\nu,l)} \epsilon_{i-j}^{(l)}, \end{aligned}$$

such that the model is a special case of (6). It is worth mentioning that the above model covers the classical case of time i idiosyncratic noise being independent across the coordinates (but possibly heteroscedastic), that is,

$$Y_{ni}^{(\nu)} = b_{n\nu}^{(1)} F_{ni}^{(1)} + \cdots + b_{n\nu}^{(L-1)} F_{ni}^{(L-1)} + \tilde{\epsilon}_{ni}^{(\nu)}$$

with $\tilde{\epsilon}_{ni}^{(1)}, \dots, \tilde{\epsilon}_{ni}^{(d_n)}$ independent (namely for $c_{nj}^{(\nu)} = 0$, if $0 \leq j \neq \nu$, and $\neq 0, j = \nu$), but also allows for time i correlated noise terms, which is a common and realistic assumption to handle the case that the factors do not absorb all correlations.

A special case well known from the literature is to assume that $\boldsymbol{Y}_{ni} = \mathbf{B}_n \mathbf{F}_{ni} + \mathbf{E}_{ni}$ as above with a VAR model for the factor process,

$$\mathbf{F}_{ni} = \Phi(L) \mathbf{F}_{n,i-1} + \mathbf{U}_{ni},$$

with mean zero errors and a lag polynomial $\Phi(L)$ of finite order such that $\det(\mathbf{I} - \Phi(z)) = 0$ has all its roots outside the unit circle. If $\mathbf{B}_n$ has full rank and if the two innovation series form a bivariate weak white noise which is orthogonal to all lagged factor values, then $\mathbf{Y}_{ni}$ attains a VARMA representation, see Dufour and Stevanovic (2013). A similar result holds for VARMA factors. Our framework and discussion go beyond these settings, since general linear processes with slowly decaying coefficients are considered and, as shown by the following example, models with an infinite number of factors are covered.

Example 2 (A generalized factor model). In some applications the assumption of independent factors is too restrictive. Our model framework (3) allows for the more general case of $K + K'$ factors, where K factors are as above and thus guaranteed to be independent. The additional K' factors may be correlated among each other but independent of the other factors. But, under certain conditions, they can be independent among each other as well. Indeed, we may even consider the case of an infinite number of correlated factors. Observe that in Example 1 all $K(=L-1)$ factors and the errors depend on different innovation processes, such that they are independent from each other. We now introduce, in addition to (5), further factors depending on the same innovation process $\{\varepsilon_i^{(L-1)} : i \geq 0\}$, by defining

$$F_{ni}^{(l)} = \sum_{j=0}^{\infty} g_{nj}^{(l)} \varepsilon_{i-j}^{(L-1)}, \quad \text{for } l = L, \dots, L-1+K',$$

for coefficients $\{g_{nj}^{(l)} : j \geq 0\}$. In general, these factors are correlated, but they may be independent among each other as well as independent of the other K factors, namely if the sets $\mathcal{I}_{n\ell} = \{j \geq 0 : g_{nj}^{(l)} \neq 0\}$ are disjoint (for each n). We obtain

$$\begin{aligned} Y_{ni}^{(\nu)} &= \sum_{l=1}^{K+K'} b_{n\nu}^{(l)} F_{ni}^{(l)} + E_{ni}^{(\nu)} = \sum_{l=1}^{L-2} b_{n\nu}^{(l)} F_{ni}^{(l)} + \sum_{l=L-1}^{K+K'} b_{n\nu}^{(l)} F_{ni}^{(l)} + E_{ni}^{(\nu)} \\ &= \sum_{l=1}^{L-2} \sum_{j=0}^{\infty} \varepsilon_{i-j}^{(l)} \underbrace{b_{n\nu}^{(l)} g_{nj}^{(l)}}_{=: c_{nj}^{(\nu,l)}} + \sum_{j=0}^{\infty} \varepsilon_{i-j}^{(L-1)} \underbrace{\sum_{l=L-1}^{K+K'} b_{n\nu}^{(l)} g_{nj}^{(l)}}_{=: c_{nj}^{(\nu,L-1)}} + \sum_{j=0}^{\infty} \underbrace{c_{nj}^{(\nu)}}_{=: c_{nj}^{(\nu,L)}} \varepsilon_{i-j}^{(L)} \\ &= \sum_{l=1}^{L} \sum_{j=0}^{\infty} c_{nj}^{(\nu,l)} \varepsilon_{i-j}^{(l)}. \end{aligned}$$

Of course, all sequences of coefficients have to satisfy Assumption 3. Observe that, in view of the representation in the second line in the above display, K' and hence K can be even ∞, provided we impose sufficient conditions to ensure that the coefficients $c_{nj}^{(\nu,L-1)}$ decay fast enough. A sufficient condition on the factor loadings $b_{ni}^{(l)}$, for $i = 1, \dots, d, l = 1, \dots, K+K'$, is to assume that

$$\sup_{n\in\mathbb{N}} \max_{1\leq i\leq d_n} |b_{ni}^{(l)}|^2 \ll l^{-2-\delta}, \quad \text{for some } \delta > 0.$$

3 | CHANGE-POINT TESTS FOR TESTING THE STABILITY OF THE COVARIANCE STRUCTURE AND FOR FACTOR MODELS

Change-point testing based on CUSUM statistics is a widespread approach to test for the presence of changes. Here we study several CUSUM procedures and a specialized approach to handle factor models. Estimation of the change-point is studied as well.

3.1 | Testing the stability of the covariance structure

Suppose we are interested in testing whether there is a change in the covariance matrix of the vector time series $\mathbf{Y}_{n1}, \dots, \mathbf{Y}_{nn}$. Thus, let us assume that

$$\mathbf{Y}_{ni} = \mathbf{Y}_{ni}^{(0)} \mathbb{1}(i \leq \tau) + \mathbf{Y}_{ni}^{(1)} \mathbb{1}(i > \tau), \quad 1 \leq i \leq n,$$

for some change-point $1 \leq \tau \leq n$, where $\mathbf{Y}_{ni}^{(0)}$ denotes the stationary time series before the change and $\mathbf{Y}_{ni}^{(1)}$ the stationary after-change time series. If $\tau < n$, then the change occurs within the observed sample. The change may be due to a change of the variance of the innovations or because of a change of the coefficients of the assumed linear process framework. In any case, we assume that it induces a change of the covariance matrix from

$$\mathbf{\Sigma}_n^{(0)} = \mathrm{Var}(Y_{n1}^{(0)})$$

to a different covariance matrix

$$\mathbf{\Sigma}_n^{(1)} = \mathrm{Var}(Y_{n,\tau+1}^{(1)}).$$

That means, denoting $\mathbf{\Sigma}_n[i] = \mathrm{Var}(\mathbf{Y}_{ni})$, $1 \leq i \leq n$, we are interested in the change-point testing problem

$$\mathrm{H}_0 \colon \mathbf{\Sigma}_n[i] = \mathbf{\Sigma}_n^{(0)} \quad \forall i = 1, \dots, n \qquad \mathrm{H}_1 \colon \exists\, k \in \{2, \dots, n\} : \mathbf{\Sigma}_n[k] \neq \mathbf{\Sigma}_n^{(0)},$$

which tests the null hypothesis of stability of the covariance structure against the alternative hypothesis of a change, equivalently expressed as

$$\mathrm{H}_0 \colon \tau = n \quad \mathrm{H}_1 \colon \tau < n.$$

Let $\mathcal{V}_n = \{(\mathbf{x}_n, \mathbf{y}_n) \in \mathbb{R}^{d_n} \times \mathbb{R}^{d_n} : \mathbf{x}_n' \mathbf{\Sigma}_{n0} \mathbf{y}_n \neq \mathbf{x}_n' \mathbf{\Sigma}_{n1} \mathbf{y}_n\}$. Select weighting vectors $\mathbf{v}_n, \mathbf{w}_n \in \mathbf{V}_n$ satisfying Assumption 1 (i) resp. (ii). Then the change is also present in the sequence of bilinear forms

$$\sigma_n^2[k] = \mathbf{v}_n' \mathbf{\Sigma}_n[k] \mathbf{w}_n, \quad k \geq 1.$$

We assume that

$$\inf_{n \geq 1} |\Delta_n| > 0, \quad \Delta_n = \mathbf{v}_n' \mathbf{\Sigma}_n^{(0)} \mathbf{w}_n - \mathbf{v}_n' \mathbf{\Sigma}_n^{(1)} \mathbf{w}_n, \tag{7}$$

to ensure that the change is present asymptotically. The proposed change-point tests are based on the partial sums of outer products,

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

The maximally selected CUSUM statistic compares $\mathbf{S}_{nk}/k$, the sample covariance matrix using the data available at time k, with the full-sample estimate in terms of the bilinear form induced by the weighting vectors. Precisely, it is defined as

$$C_n = T_n / \hat{\alpha}_n \quad \text{with} \quad T_n = \max_{1 \leq k < n} \frac{1}{\sqrt{n}} \left| \mathbf{v}_n' \left(\mathbf{S}_{nk} - \frac{k}{n} \mathbf{S}_{nn} \right) \mathbf{w}_n \right|. \tag{8}$$

Here $\hat{\alpha}_n$ is an estimator of the asymptotic standard deviation associated to the bilinear form under study and is discussed in detail below. The theoretical results provided in the next section yield the asymptotic distribution under the null hypothesis, justifying the following test: Reject H_0 at the nominal significance level $\alpha \in (0,1)$, if $C_n > c_{1-\alpha}$, where c_p denotes the p-quantile of the Kolmogorov distribution, see (13). If the test rejects, it is of interest to locate the change. Under the common assumption that the true change-point τ is proportional to the sample size, $\tau = \lfloor n\vartheta \rfloor$, $0 < \varepsilon < \vartheta < 1$, one may estimate it consistently, as shown in the next section, by

$$\hat{\tau}_n = \underset{n_0 \le k < n}{\arg\max} \frac{1}{n} \left| \mathbf{v}_n' \left(\mathbf{S}_{nk} - \frac{k}{n} \mathbf{S}_{nn} \right) \mathbf{w}_n \right|,$$

that is as the (smallest) time index $k \ge n_0 := \lfloor n\varepsilon \rfloor$, for some small $\varepsilon > 0$, at which the maximum in the definition of C_n resp. T_n is attained. Practitioners could find it more intuitive to estimate the change-point by maximizing the distance between the sample averages, that is, to use

$$\tilde{\tau}_n = \underset{n_0 \le k < n - n_0}{\arg\max} \left| \mathbf{v}_n' \left(\frac{1}{k} \mathbf{S}_{nk} \right) \mathbf{w}_n - \mathbf{v}_n' \left(\frac{1}{n-k} [\mathbf{S}_{nn} - \mathbf{S}_{nk}] \right) \mathbf{w}_n \right|.$$

To avoid estimation of the asymptotic variance parameter it has been proposed to use self-normalized CUSUMs, see Shao and Zhang (2010) and the discussion in the Introduction. Here one normalizes the CUSUM statistic, which can be represented as a functional of the partial sum process, by an appropriate integral functional of the partial sum process whose distributional limit is linear in the asymptotic variance parameter. As a consequence, this nuisance parameter cancels in the limit. Adopted to our situation here one may calculate

$$T_{n,sn} = \max_{1 \le k < n} \left| \mathbf{v}_n' \left(\mathbf{S}_{nk} - \frac{k}{n} \mathbf{S}_{nn} \right) \mathbf{w}_n \right) / V_n(k),$$

with $V_n(k) = n^{-2}[V_{n1}(k) + V_{n2}(k)]$ where

$$V_{n1}(k) = \sum_{i=1}^{k} \left\{ \mathbf{v}_n' \left(\sum_{\ell=1}^{i} \mathbf{S}_{n\ell} - \frac{i}{k} \sum_{\ell=1}^{k} \mathbf{S}_{n\ell} \right) \mathbf{w}_n \right\}^2,$$

$$V_{n2}(k) = \sum_{i=k+1}^{n} \left\{ \mathbf{v}_n' \left(\sum_{\ell=i}^{n} \mathbf{S}_{n\ell} - \frac{n-i+1}{n-k} \sum_{\ell=k+1}^{n} \mathbf{S}_{n\ell} \right) \mathbf{w}_n \right\}^2.$$

Here, as in Shao and Zhang (2010), the integrated squares of bilinear forms of CUSUMs centered at the sample average are calculated separately for the data up to and after the candidate location k for the change-point. The limiting distribution of $T_{n,sn}$ is given by law of

$$\mathcal{T}_{sn} = \sup_{0<r<1} \frac{|B(r)\text{-}rB(1)|}{\sqrt{\int_0^r [B(s) - \frac{s}{r}B(r)]^2 \, ds + \int_r^1 [B(1) - B(s) - \frac{1-s}{1-r}(B(1) - B(r))]^2 \, ds}}. \tag{9}$$

Quantiles have to be determined by simulation, however, and have been tabulated in Shao and Zhang (2010). We shall also need p values. Here one can simulate S replicates, say, $T_{1,sn}, \dots, T_{S,sn} \sim \mathcal{T}_{sn}$, of (9) once and estimate the p value of an observed statistic $T_{obs,sn}$ by $S^{-1} \sum_{s=1}^{S} \mathbf{1}(T_{s,sn} > T_{obs,sn})$.

The above CUSUM statistics are known to be powerful for a change in the middle of the sample. For an early or late change the performance gets worse, see, for example, Steland (2020), and may deteriorate completely, if there is a more complex situation with several change-points, see the simulations in Section 4. Having in mind the latter case an interesting further change-point statistic is

$$C_{n,md} = T_{n,md}/\hat{\alpha}_n \quad \text{with} \quad T_{n,md} = \max_{1\le i<j\le n} \frac{1}{\sqrt{n}} \left| \sum_{i<\ell\le j} \mathbf{v}_n' \left(\mathbf{S}_{n\ell} - \frac{(j-i)}{n}\mathbf{S}_{nn} \right) \mathbf{w}_n \right|,$$

which considers the maximal deviation from the average of the cumulated sums, the maximum being taken over all possible subsamples $\mathbf{Y}_{n\ell}, i < \ell \le j$ for $1 \le i < j \le n$. In other words, each subsample gives rise to a CUSUM statistic and the maximal one is taken. In this way, even small segments where the null hypothesis is violated can be detected, wherever they are located. We shall show that

$$T_{n,md} \xrightarrow{d} \sup_{0<s<t<1} |B^0(s) - B^0(t)|,$$

as $n \to \infty$, for some Brownian bridge B^0 on $[0, 1]$. The limiting process is the *maximal loss* functional of Brownian bridge. Its exceedance asymptotics has been studied by Dębicki et al. (2016) and satisfies

$$\mathrm{P}\left(\sup_{0<s<t<1} |B^0(s) - B^0(t)| \ge u \right) = 2^{5/2}\sqrt{\pi}u^3(1 - \Phi(2u)(1 + o(1)), \quad u \to \infty. \tag{10}$$

Quantiles are therefore approximated by solving $2^{5/2}\sqrt{\pi}u^3(1 - \Phi(2u) = \alpha$ for sufficiently small nominal significance levels α.

When being interested in testing k pairs of weighting vectors, we propose to combine the above test with a multiple testing procedure. Using our large sample approximations one may easily calculate corresponding p values p_i, $1 \le i \le k$, and reject the global null hypothesis on the global significance level $\alpha \in (0, 1)$, if $p_{(1)} \le \alpha/k$. Here we denote by $p_{(1)} \le \ldots \le p_{(k)}$ the ordered p values and by $H_0^{(i)}$ the corresponding null hypothesis of a stable pair. Furthermore, using the Benjamini–Hochberg procedure to control the familywise error rate, the null hypothesis $H_0^{(i)}$ is rejected if $p_{(i)} \le \alpha/i$. If k is large, it may be more appropriate to control the false discovery rate (FDR) by the Benjamini–Yekutieli procedure and therefore to reject $H_0^{(i)}$ if $p_{(i)} \le \alpha i/kc_i$, where $c_i = \sum_{j=1}^{i} 1/j$ ensures the FDR control under arbitrary dependence at α, if all null hypotheses are true, and at the level $\alpha k_0/k$, if k_0 null hypotheses are true, see Benjamini and Yekutieli (2001). As well known, both procedures allow for arbitrary dependencies. Our simulations indicate that multiple testing can yield substantially more powerful tests than using a only a single projection.

Depending on the problem of interest, it may be necessary to estimate the weighting vectors. Estimation of such unknowns in a procedure is a common issue and, when discussed, typically solved by (i) in-sample estimation, (ii) assuming a so-called noncontamination period as in Chu et al. (1996), or (iii) the availability of a learning sample of size m of the same dimension, such that especially $d_m = d_n$ holds. Of course, this can be ensured by splitting a given time series of length $N > n + m$ of dimension $\tilde{d}_N$ in two segments $\mathbf{Y}_{N1}, \ldots, \mathbf{Y}_{Nm}$ (the learning sample of size m) and $\mathbf{Y}_{N,N-n+1}, \mathbf{Y}_{NN}$ (the test sample of size n). Then both samples have the same dimension $d_n := \tilde{d}_N$; the condition $N > n + m$ ensures a gap which should be chosen so that the samples can be assumed to be (approximately) independent.

3.2 | Testing the stability of a high-dimensional factor model

Let us now elaborate specifically on testing the stability of a high-dimensional factor model. Assume that the learning and testing series are weakly stationary and satisfy the standing assumptions of the article. As above, denote the pre- and postchange covariance matrices by $\mathbf{\Sigma}_n^{(0)}$ and $\mathbf{\Sigma}_n^{(1)}$. Furthermore, suppose that a high-dimensional factor model holds true for the learning sample und until the change in the testing sample, such that using the notation introduced above

$$\mathbf{\Sigma}_n^{(0)} = \mathbf{B}_n \mathrm{Cov}(\mathbf{F})\mathbf{B}_n' + \mathbf{\Sigma}_{nE},$$

where $\mathbf{\Sigma}_{nE} = \mathrm{Cov}(\mathbf{E}_n)$. To fix ideas, assume a nonapproximate factor model with $\mathrm{Cov}(\mathbf{E}_n) = \sigma_E^2\mathbf{I}$. If $(\mathbf{u}, \lambda)$ is an eigenvector–eigenvalue pair of $\mathbf{B}_n\mathrm{Cov}(\mathbf{F})\mathbf{B}_n'$, then $(\mathbf{u}, \lambda + \sigma_E^2)$ is an eigenvector–eigenvalue pair of $\mathbf{\Sigma}_n^{(0)}$. This means, changes in the factors or loadings change the eigenstructure of the covariance matrix and for approximative factor models this holds in an approximative sense. Furthermore, if $\mathbf{Y}$ is a mean zero vector with covariance matrix $\mathbf{\Sigma}_n^{(0)}$ (a generic prechange observation), then $Q(\mathbf{u}) = \mathrm{E}(\mathbf{u}'\mathbf{Y}\mathbf{Y}'\mathbf{u}) = \lambda + \sigma_E^2$. If a change to $\mathbf{\Sigma}_n^{(1)}$ leaves the eigenvectors invariant but changes the eigenvalue λ to $\lambda + \delta$ for some $\delta > 0$, the quadratic form gives $Q(\mathbf{u}) = \lambda + \sigma_E^2 + \delta$, of course. On the other hand, if the eigenvectors change, say, to $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_{d_n}$, then we have $\mathbf{u} = \tilde{\mathbf{U}}\mathbf{a}$ for some $\mathbf{a} \in \mathbb{R}^{d_n}$, where $\tilde{\mathbf{U}} = (\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_{d_n})$ is orthogonal. We obtain for a generic observation $\mathbf{Y}$ after the change, that is, with mean zero and covariance matrix $\mathbf{\Sigma}_n^{(1)}$,

$$Q(\mathbf{u}) = \mathrm{E}(\mathbf{u}'\mathbf{Y}\mathbf{Y}'\mathbf{u}) = \mathbf{a}'\tilde{\mathbf{\Lambda}}\mathbf{a} = \sum_{i=1}^{d_n} \tilde{\lambda}_i a_i^2,$$

where $\tilde{\mathbf{\Lambda}} = \mathrm{diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{d_n})'$ and $\tilde{\lambda}_1, \dots, \tilde{\lambda}_{d_n}$ are the eigenvalues corresponding to $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_{d_n}$. Hence, the change is visible in the quadratic form $Q(\mathbf{u})$, provided $\sum_{i=1}^{d_n} \tilde{\lambda}_i a_i^2 \neq \lambda + \sigma_E^2$. These considerations show that eigenvectors can be used as weighting vectors to detect changes in a factor model.

Generally, if the factor model is unstable such that $\mathbf{\Sigma}_n^{(0)}$ changes to $\mathbf{\Sigma}_n^{(1)}$, the direction of the change $\mathbf{\Delta}_n = \mathbf{\Sigma}_n^{(1)} - \mathbf{\Sigma}_n^{(0)}$ matters, see also the discussion in Steland (2020) on its impact on the asymptotics of the bilinear forms used here for inference under such a change-point alternative model. Exploiting this relationship, however, requires knowledge or prior expectations about the direction of change in terms of the column space of $\mathbf{\Delta}_n$, which is rarely available. Therefore we focus on instability leading to a change visible in (the column space of) $\mathbf{\Sigma}_n^{(0)}$.

A classical factor model for $\mathbf{\Sigma}_n^{(0)}$ assuming uncorrelated factors $F_1, \dots, F_K$, uncorrelated errors $E_1, \dots, E_K$ and orthogonal columns $\mathbf{b}_{n1}, \dots, \mathbf{b}_{n,K}$ of the loadings matrix $\mathbf{B}$ can be interpreted as a series expansion with remainder term:

$$\mathbf{\Sigma}_n^{(0)} = \sum_{j=1}^{K} \mathrm{Var}(F_j)\mathbf{b}_{nj}\mathbf{b}_{nj}' + \mathbf{\Sigma}_{nE}.$$

The first K terms are a series with respect to the spanning set $\mathbf{b}_{n1}\mathbf{b}_{n1}', \dots, \mathbf{b}_{n,K}\mathbf{b}_{n,K}'$ and the covariance matrix of the idiosyncratic noise term represents the remainder term.

Any reasonable estimator $\hat{\mathbf{\Sigma}}_n^{(0)}$ of $\mathbf{\Sigma}_n^{(0)}$ being consistent under certain regularity conditions will then be close to that expansion as well. Especially, in an idealized factor model $\mathbf{B}_n\mathrm{Cov}(\mathbf{F})\mathbf{B}_n'$

dominates and $\boldsymbol{\Sigma}_{nE}$ is negligible or at least small, and thus the estimator $\hat{\boldsymbol{\Sigma}}_n^{(0)}$ should be close to $\sum_{j=1}^{K} \mathrm{Var}(F_j)\mathbf{b}_{nj}\mathbf{b}_{nj}'$. These considerations suggest to assume that $\hat{\boldsymbol{\Sigma}}_n^{(0)}$ attains a representation

$$\hat{\boldsymbol{\Sigma}}_n^{(0)} = \sum_{j=1}^{d_n} \lambda_{nj}^{(0)} \boldsymbol{\xi}_{nj}\boldsymbol{\xi}_{nj}', \tag{11}$$

where $\lambda_{n1}^{(0)} \geq \lambda_{n2}^{(0)} \geq \dots \geq \lambda_{n,d_n}^{(0)} \geq 0$ are ordered real numbers, called *pseudo-eigenvalues* in what follows, and $\boldsymbol{\xi}_{nj}$ are vectors with $\|\boldsymbol{\xi}_{nj}\|_2 = 1$, not necessarily orthogonal, which we may call *pseudo-eigenvectors*. Obviously, this assumption covers the case of the usual spectral representation with respect to the eigenvectors, but it is more general. In some cases the vectors $\boldsymbol{\xi}_{nj}$ result from the estimator's definition and satisfy sparsity constraints by its construction. Otherwise, we take the basis for granted, for example, by applying a sparse principal component analysis, see Zou et al. (2006), Witten et al. (2009), Benjamin Erichson et al. (2020), and Steland (2020) leading to such a representation with sparse directions. Let us assume that the pseudo-eigenvectors corresponding to the leading K pseudo-eigenvalues satisfy

$$\sup_{n\geq 1} \|\boldsymbol{\xi}_{nj}\|_{\ell_r} \leq C, \quad j = 1, \dots, K, \tag{12}$$

for some constant $C < \infty$, for $r = 1$ or $r = 2$, so that Assumption 1 (i) resp. (ii) holds. These considerations suggest to apply the change-point test discussed above to the test sample using the pseudo-eigenvectors $\boldsymbol{\xi}_{nj}$ as weighting vectors obtained from a learning sample. Then the change-point tests check whether or not there is a change in the pseudo-eigenvalues or pseudo-eigenvectors. As checking the stability of the leading eigenvector–eigenvalue pair is often of most interest, one may use the leading pseudo-eigenvector to conduct a stability test. One may also average the first k leading eigenvectors to check the stability of the leading part of the eigenstructure, but in our simulations this approach did not improve upon selecting only the leading one. Instead, one can use the multiple testing procedures discussed above to analyze the pseudo-eigenstructure in detail by selecting the first k pseudo-eigenvectors.

The estimator $\hat{\boldsymbol{\Sigma}}_n^{(0)}$ used in the pseudo-spectral representation can be the sample covariance matrix, a shrinkage estimator or some other regularized estimator, as discussed to some extent in the Introduction. Of course, one can also make use of the assumed factor model for that estimation, but the proposed testing methodology does not necessarily require to estimate a factor model. One can even rely on random projections as demonstrated in the simulations.

4 | ASYMPTOTIC RESULTS

This section provides large sample approximations, in terms of strong approximations, of the class of CUSUM statistics of interest and, more generally, of partial sums associated to bilinear forms of the sample covariance matrix. These results also comprise a multivariate approximation and a theorem about the CUSUM when maximized over all subsamples. Furthermore, we propose estimators for the asymptotic variances and covariances arising in these approximations and show that they are L_1-consistent uniformly in the dimension. Finally, we establish the ratio consistency of the proposed change-point estimator assuming that a change of the underlying coefficients of the vector time series induces the change in the covariance structure.

4.1 Strong approximations and related results

Recall that a strong approximation or strong invariance principle refers to results allowing to approximate a partial sum $S(n) = \sum_{i \leq n} X_i$ of mean zero random variables (or the interpolated version) by a Brownian motion, in the sense that one can redefine $\{X_i : i \geq 1\}$ on a new probability space together with a Brownian motion B, such that $|S(n) - B(n)|$ attains at least a LIL-type bound, $S(t) - B(t) = o(t^{1/2} \log(t)), t \geq 1, a.s.$, as first established by Strassen (1965) for independent random variables and martingales. Tight bounds such as $O(t^{1/2-\lambda})$, for some $\lambda > 0$, hold under moment conditions for an i.i.d. sequence, see the KMT construction of Komlós et al. (1976, theorem 2), and have been also established under dependence conditions, see, for example, Eberlein (1986), Philipp (1986), and Berkes et al. (2014), among others. Note that a strong approximation yielding $|S(t) - B(t)| = O(t^{1/2-\lambda})$ a.s., allows us to approximate the partial sum process $n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} X_i$, $t \in [0, 1]$, by a Brownian motion, also implying the central limit theorem and Donsker's invariance principle. In our treatment, we rely on a general strong approximation result of Philipp (1986) for weakly dependent random elements attaining values in a Hilbert space, which provides in our case the convergence rate $O(t^{1/2-\lambda})$ when approximating partial sums.

Let us first consider the maximally selected CUSUM statistic,

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

In order to devise the associated change-point test, we need the asymptotic null distribution of T_n, which is provided by the following result.

Theorem 1. *(i) Let $\mathbf{v}_n, \mathbf{w}_n \in \mathcal{V}_n$ be weighting vectors which satisfy Assumption 1 (i). Suppose $\mathbf{Y}_{ni}, 1 \leq i \leq n, n \geq 1$, is a vector time series satisfying the linear process resp. factor model framework (3) and Assumptions 2 and 3.*

Then, for each $n \in \mathbb{N}$, there exists an equivalent version of $T_n(\mathbf{v}_n, \mathbf{w}_n)$, again denoted by $T_n(\mathbf{v}_n, \mathbf{w}_n)$, and a standard Brownian motion $\{B_n(t) = B_n(t; \mathbf{v}_n, \mathbf{w}_n) : t \geq 0\}$, both defined on some richer probability space $(\Omega_n, \mathcal{F}_n, P_n)$, such that for some asymptotic variance parameter $\alpha_n(\mathbf{v}_n, \mathbf{w}_n)$

$$\left| T_n(\mathbf{v}_n, \mathbf{w}_n) - \alpha_n(\mathbf{v}_n, \mathbf{w}_n) \max_{k \leq n} \left| B_n^0 \left(\frac{k}{n} \right) \right| \right| = o(1),$$

a.s., as $n \to \infty$, where $B_n^{(0)}(t), t \geq 0$, is a standard Brownian bridge given by $B_n^0(t) = \overline{B}_n(t) - \frac{t}{n} \overline{B}_n(n), t \in [0, n]$, and $\overline{B}_n(s) := n^{-1/2} B_n(sn), s \in [0, 1]$, is the rescaled version of the process B_n on $[0, 1]$.

(ii) If $\mathbf{v}_n, \mathbf{w}_n \in \mathcal{V}_n$ satisfy Assumption 1 (ii) and $\hat{\alpha}_n(\mathbf{v}_n, \mathbf{w}_n)$ is a homogenous and consistent estimator of $\alpha_n(\mathbf{v}_n, \mathbf{w}_n)$, then

$$\left| C_n(\mathbf{v}_n, \mathbf{w}_n) - \max_{k \leq n} \left| B_n^0 \left(\frac{k}{n} \right) \right| \right| = o(1),$$

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

Observe that the above strong approximation for C_n yields the weak convergence

$$C_n(\mathbf{v}_n, \mathbf{w}_n) \xrightarrow{d} F_{KD}, \tag{13}$$

as $n \to \infty$, to the Kolmogorov distribution F_{KD}.

The proof of Theorem 1 is actually a consequence of more general strong approximations of the centered partial sums, $D_{nk}(\mathbf{v}_n, \mathbf{w}_n)$ given by

$$D_{nk}(\mathbf{v}_n, \mathbf{w}_n) = \mathbf{v}_n'(\boldsymbol{S}_{nk} - \mathrm{E}(\boldsymbol{S}_{nk}))\mathbf{w}_n, \quad n, k \geq 1.$$

Furthermore, define the associated càdlàg processes

$$\mathcal{D}_n(t; \mathbf{v}_n, \mathbf{w}_n) = \mathbf{v}_n' \frac{1}{\sqrt{n}}(\boldsymbol{S}_{n,\lfloor nt \rfloor} - \mathrm{E}(\boldsymbol{S}_{n,\lfloor nt \rfloor}))\mathbf{w}_n \tag{14}$$

and

$$\mathcal{D}_n^0(t, \mathbf{v}_n, \mathbf{w}_n) = \mathcal{D}_n\left(\frac{\lfloor nt \rfloor}{n}\right) - \frac{\lfloor nt \rfloor}{n}\mathcal{D}_n(1), \tag{15}$$

for $t \in [0, 1]$ and $n \geq 1$.

Theorem 2. *(i) Under the assumptions and construction of Theorem 1 (i) we have*

$$|D_{nt}(\mathbf{v}_n, \mathbf{w}_n) - \alpha_n(\mathbf{v}_n, \mathbf{w}_n)B_n(t)| \leq C_n\, n^{\frac{1}{2}-\lambda}, \quad \forall t > 0 \ a.s. \tag{16}$$

If $C_n n^{-\lambda} = o(1)$, as $n \to \infty$, this implies the strong approximation

$$\sup_{t\in[0,1]} \left| \mathcal{D}_n(t; \mathbf{v}_n, \mathbf{w}_n) - \alpha_n(\mathbf{v}_n, \mathbf{w}_n)B_n\left(\frac{\lfloor nt \rfloor}{n}\right)\right| = o(1), \quad a.s., \tag{17}$$

$$\sup_{t\in[0,1]} \left| \mathcal{D}_n^0(t; \mathbf{v}_n, \mathbf{w}_n) - \alpha_n(\mathbf{v}_n, \mathbf{w}_n)B_n^0\left(\frac{\lfloor nt \rfloor}{n}\right)\right| = o(1), \quad a.s., \tag{18}$$

as $n \to \infty$, for the rescaled version of B_n on $[0, 1]$ and $B_n^0(t) = B_n(t) - tB_n(1)$, $t \in [0, 1]$, which is a standard Brownian bridge on $[0, 1]$.

(ii) Under the assumptions of Theorem 1 (ii) and if $\lim_{n\to\infty}\alpha_n(\mathbf{v}_n, \mathbf{w}_n) > 0$ and $\alpha_n(\mathbf{v}_n, \mathbf{w}_n)$ is estimated by a homogenous and consistent estimator $\hat{\alpha}_n(\mathbf{v}_n, \mathbf{w}_n)$, then

$$\sup_{t\in[0,1]} \left| \mathcal{D}_n(t; \mathbf{v}_n, \mathbf{w}_n)/\hat{\alpha}_n(\mathbf{v}_n, \mathbf{w}_n) - B_n\left(\frac{\lfloor nt \rfloor}{n}\right)\right| = o(1), \quad a.s., \tag{19}$$

$$\sup_{t\in[0,1]} \left| \mathcal{D}_n^0(t; \mathbf{v}_n, \mathbf{w}_n)/\hat{\alpha}_n(\mathbf{v}_n, \mathbf{w}_n) - B_n^0\left(\frac{\lfloor nt \rfloor}{n}\right)\right| = o(1), \quad a.s., \tag{20}$$

as $n \to \infty$, hold.

Theorem 2 is essentially behind the asymptotic validity of all proposed tests. The following theorem establishes this in detail for the test based on $T_{n,md}$ which employs the tail asymptotics (10). In this case, which is slightly more involved than the others, one needs to consider a null sequence of significance levels to obtain a rigorous result. For the CUSUM statistic and the self-normalized CUSUM test we omit the details, since the derivations follow using similar arguments by relying on the implied FCLTs and the continuous mapping theorem, see also Steland (2020).

Theorem 3. *Assume the conditions of Theorem 2 (ii) hold. Let $p_n \in (0, 1/4)$, $n \geq 1$, be a sequence of nominal significance levels with $p_n = o(1)$, $n \to \infty$. Let u_n be the unique solution of $2^{5/2}\sqrt{\pi}u^3(1 - \Phi(2u) = p_n$. Then, under the null hypothesis of no change*

$$|\mathrm{P}(T_{n,md} > u_n) - p_n| = o(1), \quad n \to \infty.$$

The following theorem extends the previous strong approximation result to a finite number of bilinear forms with projection vectors $\mathbf{v}_{n1}, \ldots, \mathbf{v}_{nM}, \mathbf{w}_{n1}, \ldots, \mathbf{w}_{nM}$, for $M < \infty$. This is interesting in view of the proposed method to infer possible changes in a factor model structure by applying CUSUM change-point tests corresponding to sparse pseudo-eigenvectors, see the previous section, or, more generally, when correlated high-dimensional data is analyzed by projecting it onto a lower dimensional space to reduce complexity. For this purpose, (sparse) principal component analysis is a frequently chosen method, which reduces a large set of variables to a small set that still contains most of the information.

Theorem 4. *Let $\{\mathbf{v}_{nj}, \mathbf{w}_{nj} : 1 \leq j \leq M\}$ be weighting vectors of dimension d_n satisfying Assumption 1 (i). Then, under the assumptions as in Theorem 2 (i), there exists a M-dimensional Brownian motion $\{\mathbf{B}^{(n)}(t) : t \in [0, 1]\}$, such that*

$$\left\| (\mathcal{D}_n(t; \mathbf{v}_{ni}, \mathbf{w}_{ni}))_{i=1}^M - \left(B_n \left(\frac{\lfloor nt \rfloor}{n}; \mathbf{v}_{ni}, \mathbf{w}_{ni} \right) \right)_{i=1}^M \right\| = o(1), \quad a.s., \text{ as } n \to \infty,$$

where $\|\cdot\|$ denotes an arbitrary vector norm on $\mathbb{R}^M$. The mean zero Brownian motion $\{\mathbf{B}^{(n)}(t) : t \in [0, 1]\}$ with coordinates $B_{ni} = B_n(t; \mathbf{v}_{ni}, \mathbf{w}_{ni})$, $t \in [0, 1]$, $i = 1, \ldots, M$, is characterized by

$$\mathrm{E}[B_n^2(1; \mathbf{v}_{ni}, \mathbf{w}_{ni})] = \alpha_n(\mathbf{v}_{ni}, \mathbf{w}_{ni}), \quad \text{for } i = 1, \ldots, M,$$

and

$$\mathrm{E}[B_n(1; \mathbf{v}_{ni}, \mathbf{w}_{ni}) B_n(1; \mathbf{v}_{nj}, \mathbf{w}_{nj})] = \beta_n(\mathbf{v}_{ni}, \mathbf{w}_{ni}, \mathbf{v}_{nj}, \mathbf{w}_{nj}), \quad \text{for } 1 \leq i, j \leq M, \ i \neq j.$$

4.2 | Estimation of the asymptotic covariance structure

In general, the asymptotic variance and covariance parameters arising in the above theorems are unknown. In order to use our strong approximations, we have to find proper estimates, which requires somewhat stronger assumptions. So let us consider pairs $(\mathbf{v}_n, \mathbf{w}_n), (\mathbf{v}_{nr}, \mathbf{w}_{nr})$, $1 \leq r \leq M$, of projection vectors, on which we project the vector time series $\mathbf{Y}_{ni}$. We follow previous works (Steland & von Sachs, 2017, 2018) in using appropriate long-run variance (LRV) estimators and show that they are L_1-consistent for the general high-dimensional multivariate linear process framework. We further go beyond these results in allowing the projection vectors as well as the coefficients of the linear processes to depend on n.

It turns out that the asymptotic variance and covariance parameters, $\alpha_n^2 = \alpha_n^2(\mathbf{v}_n, \mathbf{w}_n)$ and $\beta_n^2(r, s) = \beta^2(\mathbf{v}_{nr}, \mathbf{w}_{nr}, \mathbf{v}_{ns}, \mathbf{w}_{ns})$, $1 \leq r, s, \leq M$, are the long-run variance and covariance parameters associated to the time series $\{\mathbf{v}_n' \mathbf{Y}_{ni} \mathbf{w}_n' \mathbf{Y}_{ni} : i = 1, \ldots, n\}$ and $\{\mathbf{v}_{nr}' \mathbf{Y}_{ni} \mathbf{w}_{ns}' \mathbf{Y}_{ni} : i = 1, \ldots, n\}$, respectively. In view of this fact, we use Bartlett-type LRV estimates in the time domain. To simplify notation, put $Y_{ni}(\mathbf{v}_r) = \mathbf{v}'_{nr}\mathbf{Y}_{ni}$ and so on. Observe that sample mean and sample autocovariances of these processes are given by

$$\hat{\mu}_n^{(\mathbf{v}_r,\mathbf{w}_r)} := \frac{1}{n}\sum_{j=1}^{n} Y_{nj}(\mathbf{v}_r)Y_{nj}(\mathbf{w}_r), \quad \text{for } 1 \leq r \leq M,$$

$$\hat{\Gamma}_n^{(r,s)}(h) := \frac{1}{n}\sum_{i=1}^{n-h}\left(Y_{ni}(\mathbf{v}_r)Y_{ni}(\mathbf{w}_r) - \hat{\mu}_n^{(\mathbf{v}_r,\mathbf{w}_r)}\right)\left(Y_{n,i+|h|}^{(\mathbf{v}_s)}Y_{n,i+|h|}^{(\mathbf{w}_s)} - \hat{\mu}_n^{(\mathbf{v}_s,\mathbf{w}_s)}\right), \quad |h| < n,$$

for $1 \leq r, s, \leq M$. For brevity of notation, let us denote $\hat{\Gamma}_n(h) = \hat{\Gamma}_n^{(1,1)}(h)$ in case that a single pair, $(\mathbf{v}_n, \mathbf{w}_n)$, as in Theorem 2, of weighting vectors is considered. Define

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

$$\hat{\beta}_n^2(r,s) = \hat{\beta}_n^2(r,s;d) = \hat{\Gamma}_n^{(r,s)}(0) + 2\sum_{h=1}^{m} w_{mh}\hat{\Gamma}_n^{(r,s)}(h), \quad 1, \leq r, s, \leq M,$$

where $m = m_n, n \geq 1$, is a sequence of lag truncation constants and w_{mh} are weights.

The following theorem provides the L_1-consistency, uniformly over the dimension d. It generalizes (Steland & von Sachs, 2017, th. 4.4) to projection vectors depending on the sample size and this fills a gap in the rigorous mathematical justification of the procedures when using uniformly ℓ_2-bounded projections.

Theorem 5. *Assume the weights $\{w_{mh} : h \in \mathbb{Z}, m \in \mathbb{N}\}$ satisfy*

(W1) $w_{mh} \to 1$, as $m \to \infty$, for all $h \in \mathbb{Z}$.
(W2) $0 \leq w_{mh} \leq W < \infty$ for some constant W, for all $m \geq 1, h \in \mathbb{Z}$.

Furthermore, suppose that the coefficients $c_{nj}^{(v,l)}$ ensure the decay condition

$$\sup_{n\geq 1}\sup_{1\leq v}|c_{nj}^{(v,l)}| \ll (j \vee 1)^{-(1+\delta)}, \quad \forall l = 1, \ldots, L,$$

for some $\delta > 0$, and that $\epsilon_k^{(l)}, k \geq 1$, are i.i.d. with $\max_k \mathrm{E}|\epsilon_k^{(l)}|^8 < \infty$, for $l = 1, \ldots, L$. Finally, suppose that $m_n \to \infty$ with $m^2/n = o(1)$, as $n \to \infty$.

(i) If $\mathbf{v}_n, \mathbf{w}_n$ are weighting vectors with $\sup_{n\geq 1}\|\mathbf{v}_n\|_{\ell_1}, \|\mathbf{w}_n\|_{\ell_1} < \infty$, then

$$\sup_{d\in\mathbb{N}} \mathrm{E}|\hat{\alpha}_n^2(d) - \alpha^2(d)| \underset{n\to\infty}{\to} 0.$$

(ii) Furthermore, if $\mathbf{v}_{nr}, \mathbf{w}_{nr}, r \geq 1$, are weighting vectors which satisfy Assumption 1 (i), that is, they have uniformly bounded ℓ_1-norm, then

$$\sup_{r,s\geq}\sup_{d\in\mathbb{N}} \mathrm{E}|\hat{\beta}_n^2(r,s;d) - \beta^2(r,s;d)| \underset{n\to\infty}{\to} 0.$$

In the above theorem the dimension is regarded as a formal variable d and consistency is shown uniformly in d. By plugging in the dimension d_n of the time series, this result yields the consistency of $\hat{\alpha}_n^2(d_n)$ for an growing dimension $d_n \to \infty$, without any constraint on the growth of d_n, since

$$\mathrm{E}|\hat{\alpha}_n^2(d_n) - \alpha^2(d_n)| \leq \sup_{d \in \mathbb{N}} \mathrm{E}|\hat{\alpha}_n^2(d) - \alpha^2(d)|,$$

and a similar statement holds true for the estimates $\hat{\beta}_n^2(r, s; d)$, $1 \leq r, s \leq M$.

4.3 | Consistency of the change-point estimator

Let us now consider the proposed change-point estimator $\hat{\tau}_n$, defined as the smallest time point where the test statistic in (8) attains its maximum over the time span from n_0 to n. Clearly, the estimator $\tilde{\tau}_n$ can be represented as

$$\tilde{\tau}_n = \underset{\lfloor n\varepsilon \rfloor \leq k \leq n}{\arg\max} |\hat{\mathcal{U}}_n(k)|,$$

where

$$\hat{\mathcal{U}}_n(k) = \frac{1}{n} U_{nk} - \frac{k}{n^2} U_{nn}, \quad 1 \leq k \leq n, n \geq 1,$$

with $U_{nk} = U_{nk}(\mathbf{v}_n, \mathbf{w}_n) = \mathbf{v}_n' \tilde{\mathbf{\Sigma}}_{nk} \mathbf{w}_n$, and $\arg\max_{x \in D} f(x)$ denotes the smallest maximizer of a function f defined on D. Alternatively, one may use

$$\tilde{\tau}_n = \underset{\lfloor n\varepsilon \rfloor \leq k \leq \lfloor n(1-\varepsilon) \rfloor}{\arg\max} |\tilde{\mathcal{V}}_n(k)|,$$

where

$$\tilde{\mathcal{V}}_n(k) = \frac{1}{k} U_{nk} - \frac{1}{n-k}(U_{nn} - U_{nk}).$$

Let us now assume that the time series before and after the change is given by two different coefficient arrays $\mathfrak{b} = \{b_{nj}^{(v,l)} : j \geq 0, v = 1, \ldots, d_n, l = 1, \ldots, L, n \geq 1\}$ and $\mathfrak{c} = \{c_{nj}^{(v,l)} : j \geq 0, v = 1, \ldots, d_n, l = 1, \ldots, L, n \geq 1\}$ ensuring that

$$\mathbf{\Sigma}_n^{(0)} = \mathrm{Var}(\mathbf{Y}_{n1}(\mathfrak{b})) \neq \mathrm{Var}(\mathbf{Y}_{n,\tau+1}(\mathfrak{c})) = \mathbf{\Sigma}_n^{(1)}, \tag{21}$$

where $\mathbf{Y}_{ni}(\mathfrak{b})$ resp. $\mathbf{Y}_{ni}(\mathfrak{c})$ denote the vector time series corresponding to the coefficient array $\mathfrak{b}$ resp. $\mathfrak{c}$. For $\mathbf{v}_n, \mathbf{w}_n \in \mathcal{V}_n$ this induces a change in the associated quadratic forms, such that $\Delta_n \neq 0$ where $\Delta_n = \mathbf{v}_n' \mathbf{\Sigma}_n^{(0)} \mathbf{w}_n - \mathbf{v}_n' \mathbf{\Sigma}_n^{(1)} \mathbf{w}_n$ in view of (21). In Steland (2020) it has been shown that

$$\mathcal{U}_n(k) = \mathrm{E}(\hat{\mathcal{U}}_{nk}) = \begin{cases} \frac{k(n-\tau)}{n} \Delta_n, & 1 \leq k \leq \tau, \\ \tau \frac{n-k}{n} \Delta_n, & \tau < k \leq n. \end{cases}$$

Furthermore, it holds

$$\mathcal{V}_n(k) = \mathrm{E}(\hat{\mathcal{V}}_n(k)) = \begin{cases} \frac{n-\tau}{n-k}\Delta_n, & 1 < k \leq \tau, \\ \frac{\tau}{k}\Delta_n, & \tau < k < n. \end{cases}$$

Observe that for $\Delta_n > 0$ the functions $\mathcal{U}_n(k)$ and $\mathcal{V}_n(k)$ have a unique maximum at $k = \tau$, are strictly increasing for $k < \tau$ and strictly decreasing for $k > \tau$. We strengthen condition (7) and assume now that

$$0 \neq \Delta = \lim_{n\to\infty} \Delta_n, \quad i = 0, 1. \tag{23}$$

Furthermore, let us specify the location of the change-point by assuming that

$$\tau = \lfloor n\vartheta \rfloor \tag{24}$$

for some change-point parameter $\vartheta \in (0, 1)$. ϑ can be estimated by

$$\hat{\vartheta}_n = \hat{\tau}_n/n \quad \text{resp.} \quad \tilde{\vartheta}_n = \tilde{\tau}_n/n.$$

Clearly, any estimator is ratio-consistent for τ, if it is consistent for $\vartheta > 0$.

Theorem 6. *Suppose that (23) and (24) hold true. Then the change-point estimators $\hat{\tau}_n$ and $\tilde{\vartheta}_n$ are weakly consistent and ratio-consistent for τ, that is, specifically,*

$$|\tilde{\vartheta}_n/\vartheta - 1| \stackrel{\mathrm{P}}{\to} 0, \quad |\tilde{\vartheta}_n/\vartheta - 1| \stackrel{\mathrm{P}}{\to} 0, \quad n \to \infty.$$

5 | SIMULATIONS AND DATA ANALYSIS

The proposed methods were investigated by simulations and applied to a real econometric data set.

5.1 | Simulations

To investigate size and power of the tests, two models employed in the literature on change detection in factor models are examined.

Model 1: Time series from the frequently used simple model for a d-dimensional vector time series

$$\mathbf{Y}_t = \varrho \mathbf{Y}_{t-1} + \mathbf{e}_t + \vartheta \mathbf{e}_{t-1},$$

were simulated as H_0 data, where $\mathbf{e}_t$ is i.i.d. $N(\mathbf{0}, \mathbf{I})$. Two types of alternatives were examined. First, a change-point in the middle of the sample where the covariance matrix of the $\mathbf{e}_t$ changes to $\Delta(1 + \sqrt{\log\log T/\sqrt{T}})\mathbf{I}$, $\Delta \geq 1$, similar as in Kao et al. (2018), who studied a specialized test for a linear mapping of eigenvalues and eigenvectors, respectively, given by a quadratic form. The results are, however, only to some extent comparable, since their test requires to invert a $d \times d$ matrix and is trimmed as well, thus requiring the condition $d/n < 1/2$ to be computable. Second, the case of several change-points was studied where it is known that the classical CUSUM statistic may not work. Here the covariance matrix of the $\mathbf{e}_t$ is given by

$$(1+2\Delta)\mathbf{I}, \quad \text{if } n/16 < t \leq n/4 \text{ or } (3/4)n < t \leq (15/16)n \tag{25}$$

and $\mathbf{I}$ otherwise. We consider the cases $d = 3, 10, 200, 500$ and model parameters $\varrho = 0, 0.5$ and $\vartheta = -0.5, 0, 0.5$, combined as given in the tables. The high-dimensional case was investigated for $d = n$, a setting which cannot be handled by most existing tests in the literature. The single tests use the leading eigenvector from the sample covariance matrix of a learning sample, whereas the multiple testing approach employs the leading $\max(3, \sqrt{d})$ eigenvectors. The size of the learning sample was chosen as $\max(200, n/2)$. The asymptotic variance parameters, $\alpha_n(\mathbf{v}_n, \mathbf{w}_n)$, were calculated using the quadratic spectral kernel with a lag truncation of $n^{0.3}$. For multiple testing the Benjamini–Yekutieli procedure was used. Each rejection rate was simulated using 10,000 runs, where all tests were calculated on the same prewhitened simulated time series. For the self-normalized CUSUM the required quantiles and p values under the limiting law (9) were simulated using 10,000 runs based on standard normal partial sums.

One can observe from Table 1 that the size of both the single and multiple test are very good, especially for the self-normalized statistic. When it comes to very high dimensions, the multiple tests based on the CUSUM and the maximal CUSUM over subsamples are conservative, whereas the self-normalized test develops a slight tendency to overreact, which is, however, not relevant for typical practical purposes.

Figure 1 depicts power curves for sample size $T = 200$, dimension $d = 10$ and model parameters $\varrho = 0.5$, $\vartheta = 0$. Under the alternative change-point model 2 the self-normalized CUSUM is slightly more powerful for small changes, but the CUSUM is better for large changes. The maximal CUSUM over subsamples is less performant. But under the complex change structure of change-point model (25) both the CUSUM and the self-normalized version completely break down. The power of the self-normalized CUSUM even becomes non-monotonic. It is here where the maximal CUSUM overs subsamples really shines. It is highly powerful in detecting the change structure.

Model 2: An explicit dynamic factor model $\mathbf{Y}_{ni} = \mathbf{B}_{ni}\mathbf{F}_{ni} + \mathbf{E}_{ni}$ was simulated using a specification as in Han and Inoue (2015), a factor model with L Gaussian factors, random factor loadings, and a break in the middle of the sample under the change alternative hypothesis parameterized by δ in a certain fraction of the coordinates, namely

$$Y_i^{(\nu)} = \sum_{\ell=1}^{L} b_\nu^{(\ell)} F_i^{(\ell)} + \kappa E_i^{(\nu)}, \quad 1 \leq i \leq n/2, \tag{26}$$

and

$$Y_i^{(\nu)} = \sum_{\ell=1}^{L} (b_\nu^{(\ell)} - \delta \mathbf{1}(\ell \leq La) F_i^{(\ell)} + \kappa E_i^{(\nu)}, \quad n/2 < i \leq n, \tag{27}$$

for $\delta \geq 0$ (with $\delta = 0$ corresponding the no change null hypothesis), $a \in [0, 1]$ (fraction of coordinates affected by a change under H_1 if $\delta > 0$) and, with $F_i^{(\ell)}, E_i^{(\nu)} \overset{iid}{\sim} N(0, 1)$,

TABLE 1 Size of CUSUM, maximal partial sum (max P-SUM), and self-normalized CUSUM (CUSUM-SN) tests for the leading eigenvector (single) and the multiple test (multi)

	Test	d	T	$(0, 0)$	$(0.5, 0)$	$(0, 0.5)$	$(-0.5, 0)$
Single	CUSUM	10	200	0.031	0.029	0.030	0.032
	Max P-SUM	10	200	0.045	0.039	0.044	0.043
	CUSUM-SN	10	200	0.051	0.048	0.053	0.053
Multi	CUSUM	10	200	0.011	0.010	0.011	0.012
	Max P-SUM	10	200	0.009	0.009	0.011	0.009
	CUSUM-SN	10	200	0.042	0.044	0.048	0.046
Single	CUSUM	10	500	0.035	0.035	0.039	0.042
	Max P-SUM	10	500	0.059	0.061	0.064	0.063
	CUSUM-SN	10	500	0.055	0.053	0.050	0.050
Multi	CUSUM	10	500	0.022	0.022	0.027	0.024
	Max P-SUM	10	500	0.028	0.029	0.035	0.026
	CUSUM-SN	10	500	0.044	0.042	0.046	0.046
Single	CUSUM	200	200	0.029	0.028	0.031	0.035
	max P-SUM	200	200	0.040	0.040	0.044	0.046
	CUSUM-SN	200	200	0.046	0.051	0.052	0.057
Multi	CUSUM	200	200	0.001	0.001	0.001	0.001
	Max P-SUM	200	200	0.001	0.001	0.001	0.001
	CUSUM-SN	200	200	0.067	0.066	0.065	0.064
Single	CUSUM	200	500	0.034	0.036	0.038	0.040
	Max P-SUM	200	500	0.059	0.058	0.060	0.063
	CUSUM-SN	200	500	0.050	0.047	0.052	0.051
Multi	CUSUM	200	500	0.007	0.008	0.009	0.010
	Max P-SUM	200	500	0.006	0.007	0.007	0.009
	CUSUM-SN	200	500	0.059	0.063	0.065	0.058

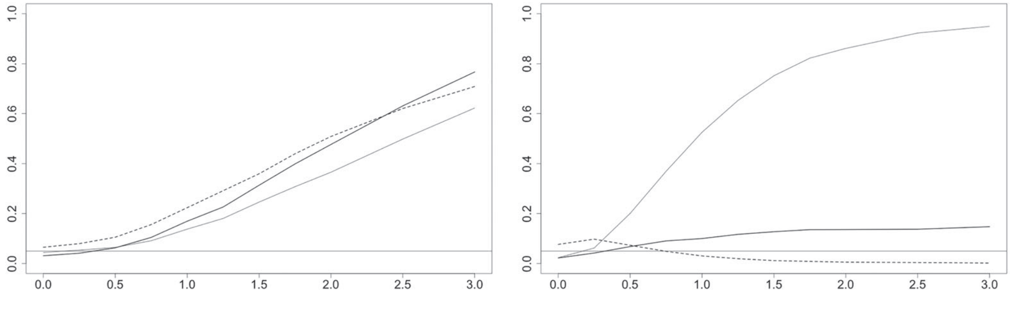

FIGURE 1 Power curves as a function of Δ for the CUSUM test (bold), self-standardized CUSUM (dashed), and the maximal CUSUM test (thin) for $\varrho = 0.5$, $\vartheta = 0$. Left panel: change in the middle of the sample. Right panel: changes as specified in (25).

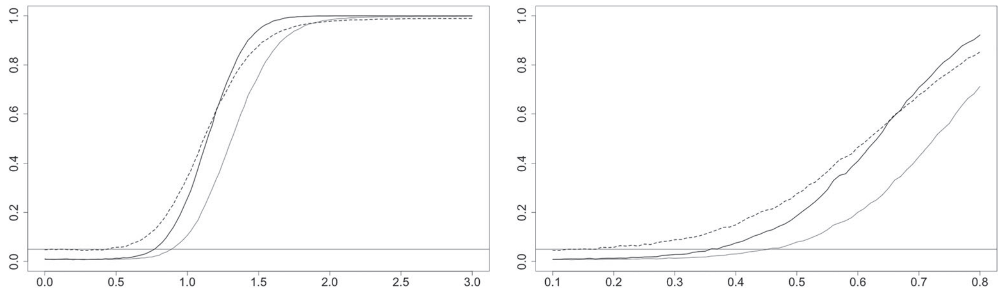

FIGURE 2 Left panel: Power curves for the dynamic factor model given by (26) and (27) as a function of δ (specifying the magnitude of the change in the middle of the sample), for the CUSUM test (bold), self-standardized CUSUM (dashed), and the maximal CUSUM test (thin). Right panel: power for $\delta = 2$ and $d = 200$ as a function of the fraction, a, of coordinates affected by a change

$\kappa = \sqrt{(1 + \delta^2/4)L}$ and random factor loadings $b_\nu^{(\ell)} \overset{iid}{\sim} N(\delta/2, 1)$. In this model the first $\lfloor La \rfloor$ coordinates of the vector time series are affected by a change in the factor loadings. In-sample estimates of the eigenvectors were used as weighting vectors. The remaining parameters were as in the first set of simulations.

For $a = 1$ and $L = 3$ factors the power curves as a function of δ, simulated for $L = 3$ on the grid $\delta = 0, 0.5, \ldots, 3$, are shown in the left panel of Figure 2. In comparison with the tests examined in Han and Inoue (2015), Wald and LM tests based on differences of pre- and post-sample means of outer products of PCA estimates of the factors maximized over the candidate location for the change, our tests have somewhat less power for small values of δ, but for larger values they compete with the best tests. Especially, the power curves are monotone, whereas several of the tests studied in the cited paper have non-monotonic power in δ with power dropping down to 0.6 or even 0.5 for $\delta = 2$. Furthermore, the right plot in Figure 2 shows the power as a function of the fraction, a, of time series under change, for $\delta = 2$ and $d = 200$. These results are more or less in agreement with the findings for model 1.

Finally, it was examined how the power of the tests is affected by the number of projections used for the multiple test. Figure 3 shows the interesting results when using estimated eigenvectors as well as random projections (i.i.d. standard normal). It turns out that the CUSUM with LRV estimation reaches its highest power for a relatively small number of projections, whereas the power of the self-standardized test is increasing. For large values, it competes well with the best tests in Han and Inoue (2015). Random projections provides very similar results for the CUSUM with estimated asymptotic variance and the maximal CUSUM, but lead to loss of power when using the self-standardized CUSUM with many projections.

5.2 | Data example: Is there an impact of the SARS-CoV-2 coronavirus pandemic on the Fama/French factor model?

To illustrate the examined change-point test we consider the well-known Fama and French five-factor model, (Fama & French, 2015), using data as provided by French. The five-factor model extends the original three-factor model, Fama and French (1993), which added size and value to an assets beta, by considering in addition factors for profitability and investment. As well known, these factors are economic ones and do not result from a statistical factor analysis, and thus it

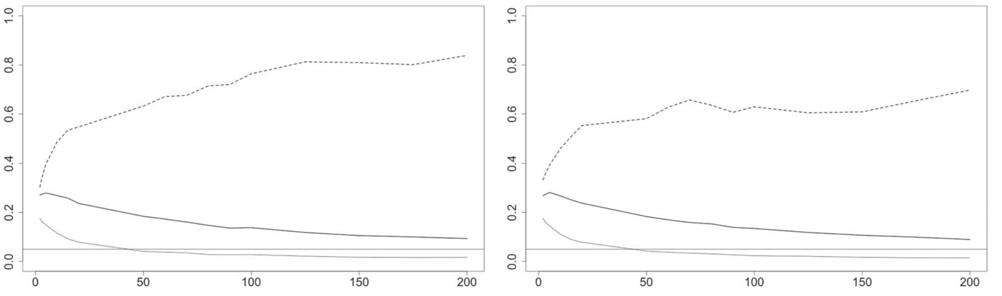

FIGURE 3 Power curves for the dynamic factor model (26) and (27) as a function of the number of projections used for multiple testing, for fixed $\delta = 2$ and a change in the middle of the sample. CUSUM test (bold), self-standardized CUSUM (dashed), and the maximal CUSUM test (thin). Left panel: results for projections onto eigenvectors. Right panel: random projections

makes sense to assume an underlying statistical factor model. The Fama and French factors are correlated and it is interesting to investigate whether for a given period the eigenstructure is affected by changes, which could be explained in the presence of a statistical factor model by a change of the statistical factors, their loadings or changes in the idiosyncratic component.

The proposed multiple testing change-point test procedure with $k = 5$ was applied to analyze the eigenstructure for three periods. Table 2 provides the p values of each test, the corresponding thresholds according to Holm's procedure to control the familywise error rate, the test decisions when using the Benjamini–Yekutieli multiple testing procedure, and the estimated change-points in case of rejection of the null hypothesis of stability of the corresponding subspace. Data from January 2019 to May 2020 was analyzed to investigate a possible break due to the SARS-CoV-2 corona virus pandemic. Furthermore, as a comparison, the year 2018 was analyzed. The data from 2015 were selected by purpose as an example illustrating a stable period.

As can be seen from Table 2, in 2015 there is no evidence in the data for a change in the eigenstructure. In the year 2018 the procedure decides in favor of a change in the leading eigenvector–eigenvalue pair. The break is dated as October 10, 2018, by both $\tilde{\tau}_n$ and $\tilde{\tau}_n$, where the losses of SP500, Dow Jones Industrial and Nasday composite were 3.3%, 3.2%, and 4.8%, and losses at other exchanges were between 2%and 6%. Market observers, such as Neue Zürcher Zeitung, argued that Donald Trump's tax cuts despite increasing deficits and the Fed's raising of benchmark rates by a quarter percentage point in September to prevent overheating of the economy could be among the major economic reasons.

The data from January 2019 until May 2020 spanning the COVID-19 coronavirus global outbreak are, however, perhaps of most interest. All tests yield very small p values and all null hypotheses are rejected indicating evidence of a substantial change in the eigenstructure. When relying on the estimator $\tilde{\tau}_n$, the estimated break dates are particularly coherent. The breaks in the eigensubspaces are dated between January 29 and February 28. Except the change in the least important (fifth) and fourth direction, all changes are indeed dated quite coherently between February 21 and February 28: On Monday, February 24, the Dow Jones Industrial Index and FTSE fell by more than 3% in view of news that the COVID-19 coronavirus outbreak spread outside China over the weekend. Other leading indices such as DAX and CAC40 followed. On February 27 the NASDAQ-100 and S&P500 indices suffered their sharpest fall since the 2008 crash.

TABLE 2 Data analysis of the Fama/French factors over three periods. BY test decision gives the results of the Benjamini–Yekutieli procedure. For the decisions of the Holm procedure compare the p-values with the Holm thresholds.

Year 2015					
p values CUSUM	0.099	0.1888	0.092	0.491	0.162
p values CUSUM-SN	0.326	0.344	0.025	0.232	0.163
p values max P-SUM	0.053	0.0253	0.056	0.225	0.042
BY Holm threshold	0.0125	0.025	0.01	0.05	0.0167
test decision (all tests)	0	0	0	0	0
Year 2018					
p value CUSUM	0.002	0.0628	0.0498	0.5246	0.6907
p values CUSUM-SN	0.216	0.059	0.171	0.611	0.55
p values max P-SUM	0.008	0.087	0.027	0.244	0.456
Holm threshold	0.0167	0.01	0.05	0.01	0.025
BY test decision CUSUM	1	0	0	0	0
BY test decision CUSUM-SN	0	0	0	0	0
BY test decision max P-SUM	1	0	0	0	0
Est. change-point ($\tilde{\tau}_n$ and $\tilde{\tau}_n$)	9.10.18	–	–	–	–
01/2019–05/2020					
p values CUSUM	$6 \cdot 10^{-4}$	$<10^{-5}$	0.003	0.01	0.025
p values CUSUM-SN	0.011	10^{-4}	0.016	0.002	0.012
p values max P-SUM	0.024	$5.8 \cdot 10^{-5}$	0.099	0.003	0.017
Holm threshold	0.0167	0.01	0.05	0.01	0.025
BY test decision CUSUM	1	1	1	1	1
BY test decision CUSUM-SN	1	1	1	1	1
BY test decision max P-SUM	0	1	0	1	0
Est. change-points ($\tilde{\tau}_n$)	21.02.20	25.02.20	28.02.20	28.01.20	29.01.20
Est. change-points ($\tilde{\tau}_n$)	28.02.20	13.03.20	09.03.20	27.01.20	28.02.20

ACKNOWLEDGMENTS

The authors thank Rainer v. Sachs, UC Louvain, for discussions and acknowledge financial support from Deutsche Forschungsgemeinschaft, grants STE 1034-11/1 and 1034-11/2. Helpful and constructive comments from anonymous reviewers are gratefully acknowledged. Open access funding enabled and organized by Projekt DEAL.

ORCID

Ansgar Steland https://orcid.org/0000-0001-9395-7458

REFERENCES

Aue, A., Hörmann, S., Horváth, L., & Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, *37*(6B), 4046–4087. https://doi.org/10.1214/09-AOS707.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, *71*(1), 135–171.

Bai, J., & Li, K. (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics*, *40*(1), 436–465.

Bai, J., & Li, K. (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension. *Review of Economics and Statistics*, *98*(2), 298–309.

Bai, J., & Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics*, *3*(2), 89–163.

Bai, J., & Shi, S. (2011). Estimating high dimensional covariance matrices and its applications. *Annals of Economics and Finance*, *12*(2), 199–215.

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*, 1281–1311.

Benjamin Erichson, N., Zheng, P., Manohar, K., Brunton, S. L., Kutz, J. N., & Aravkin, A. A. (2020). Sparse principal component analysis via variable projection. *SIAM Journal on Applied Mathematics*, *80*(2), 977–1002.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998.

Berkes, I., Liu, W., & Wei Biao, W. (2014). Komlós-Major-Tusnády approximation under dependence. *Annals of Probability*, *42*(2), 794–817. https://doi.org/10.1214/13-AOP850.

Bours, M. 2019. Large sample approximations for high-dimensional covariance matrices under factor models. *Technical Report, Institute of Statistics, RWTH Aachen University*.

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. *Journal of Econometrics*, *108*(2), 343–363. https://doi.org/10.1016/S0304-4076(01)00139-7.

Breitung, J., & Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *J. Econometrics*, *163*(1), 71–84. https://doi.org/10.1016/j.jeconom.2010.11.008.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., & Loris, I. (2009). Sparse and stable Markowitz portfolios. *PNAS*, *106*(30), 12267–12272.

Chen, C.-F. (1979). Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*(2), 235–248.

Chow, Y. S., & Teicher, H. (1997). *Probability Theory*, Springer Texts in Statistics, 3rd ed New York, NY: Springer.

Chu, C., Stinchcombe, M., & White, H. (1996). Monitoring structural change. econometrica. *Econometrica*, *64*(5), 1045–1065.

Daniels, M. J., & Kass, R. E. (1999). Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, *94*(448), 1254–1263.

Daniels, M. J., & Kass, R. E. (2001). Shrinkage estimators for covariance matrices. *Biometrics*, *57*(4), 1173–1184.

Dębicki, K., Hashorva, E., & Ji, L. (2016). Extremes of a class of nonhomogeneous Gaussian random fields. *Annals of Probability*, *44*(2), 984–1012. https://doi.org/10.1214/14-AOP994.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, *28*, 157–175.

Dufour, J.-M., & Stevanovic, D. (2013). Factor-augmented varma models with macroeconomic applications. *Journal of Business Economics Statistics*, *31*(4), 491–506. https://doi.org/10.1080/07350015.2013.818005.

Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Annals of Probability*, *14*(1), 260–270.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fan, J., Fan, Y., & Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, *147*(1), 186–197.

Fan, J., Liao, Y., & Liu, H. (2016). An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, *19*(1).C1–C32.

Fan, J., Liao, Y., & Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics*, *39*(6), 3320.

Fan, J., Liao, Y., & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, *75*(4), 603–680. https://doi.org/10.1111/rssb.12016.

Goldfarb, D., & Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, *28*(1), 1–38.

Han, X., & Inoue, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory*, *31*, 1117–1152.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, *29*(2), 295–327.

Kao, C., Trapani, L., & Urga, G. (2018). Testing for instability in covariance structures. *Bernoulli*, *24*(1), 740–771. https://doi.org/10.3150/16-BEJ894.

Kiefer, N. M., Vogelsang, T. J., & Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, *68*(3), 695–714. https://doi.org/10.1111/1468-0262.00128.

Komlós, J., Major, P., & Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. *I. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *34*(1), 33–58.

Kouritzin, M. A. (1995). Strong approximation for cross-covariances of linear variables with long-range dependence. *Stochastic Processes and Their Applications*, *60*(2), 343–353.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y.

Laloux, L., Cizeau, P., Potters, M., & Bouchaud, J.-P. (2000). Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, *3*(03), 391–397.

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, *10*(5), 603–621.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411.

Marcinkiewicz, J., & Zygmund, A. (1937). Mean values of trigonometrical polynomials. *Fundamenta Mathematicae*, *28*(1), 131–166.

Pafka, S., & Kondor, I. (2003). Noisy covariance matrices and portfolio optimization ii. *Physica A: Statistical Mechanics and Its Applications*, *319*, 487–494.

Philipp, W. (1986). A note on the almost sure approximation of weakly dependent random variables. *Monatshefte für Mathematik*, *102*(3), 227–236.

Phillips, P. C. B., & Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, *20*(2), 971–1001.

Reyment, R. A., & Jvreskog, K. G. (1996). *Applied factor analysis in the natural sciences*. Cambridge University Press.

Sancetta, A. (2008). Sample covariance shrinkage for high dimensional dependent data. *Journal of Multivariate Analysis*, *99*(5), 949–967.

Shao, X., & Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, *105*(491), 1228–1240. https://doi.org/10.1198/jasa.2010.tm10103.

Sharifi, S., Crane, M., Shamaie, A., & Ruskin, H. (2004). Random matrix theory for portfolio optimization: A stability approach. *Physica A: Statistical Mechanics and its Applications*, *335*(3), 629–643.

Stein, C. 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA. (Vol. 1, pp. 197–206).

Stein, C. 1975. Estimation of a covariance matrix. Rietz Lecture. *Proceedings of the 39th Annual Meeting IMS*. Atlanta, GA.

Steland, A. (2007). Monitoring procedures to detect unit roots and stationarity. *Econometric Theory*, *23*(6), 1108–1135. https://doi.org/10.1017/S0266466607070442.

Steland, A. (2020). Testing and estimating change-points in the covariance matrix of a high-dimensional time series. *Journal of Multivariate Analysis*, *177*, 104582.

Steland, A., & von Sachs, R. (2017). Large-sample approximations for variance-covariance matrices of high-dimensional time series. *Bernoulli*, *23*(4A), 2299–2329.

Steland, A., & von Sachs, R. (2018). Asymptotics for high–dimensional covariance matrices and quadratic forms with applications to the trace functional and shrinkage. *Stochastic Processes and their Applications*, *128*(8), 2816–2855.

Strassen V. 1965. Almost sure behaviour of sums of independent random variables and martingales. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA. (Vol. 2, pp. 315–343).

van der Vaart, A. W. (1998). *Asymptotic statistics, volume 3 of Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.

Wei Biao, W. (2007). Strong invariance principles for dependent random variables. *Annals of Probability*, *35*(6), 2294–2320.
Wei Biao, W., & Min, W. (2005). On linear processes with dependent innovations. *Stochastic Processes and Their Applications*, *115*(6), 939–958.
Witten, D. M., Tibshirani, R., & Hastie, T. (2009). A penalized decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, *10*, 515–534.
Yang, R., & Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. *Annals of Statistics*, *22*(3), 1195–1211.
Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, *15*(2), 265–286. https://doi.org/10.1198/106186006X113430.

How to cite this article: Bours M, Steland A. Large-sample approximations and change testing for high-dimensional covariance matrices of multivariate linear time series and factor models. *Scand J Statist*. 2021;1–45. https://doi.org/10.1111/sjos.12508

APPENDIX. PROOFS

This section provides proofs of several main results. Most omitted derivations and details as well as additional material can be found for sections A.1–A.4 in Bours (2019) and Steland (2020). The general method of proof is as in previous works, especially Steland and von Sachs (2017), but the structure of the approximating martingale is considerably more involved for the model studied here and requires its own treatment.

Following previous developments (see, e.g., Kouritzin, 1995; Phillips & Solo, 1992; Steland & von Sachs, 2017), we show a strong approximation for D_{nk} by using results from Philipp (1986). The basic idea of the proof is to approximate the bilinear form D_{nk} by a martingale array.

A.1 Preliminaries and additional results

Multivariate time series with colored noise are covered by model (3) due to Lemma 1.

Proof of Lemma 1. We may assume that $\mathbf{G}$ has rank L. Let $\mathbf{G} = \sum_{k=1}^{L} \varrho_k \mathbf{u}_k \mathbf{v}_k'$ denote the singular value decomposition of $\mathbf{G}$ with left-singular vectors $\mathbf{u}_k = (u_{k1}, \dots, u_{k,d_n})'$, right-singular vectors $\mathbf{v}_k = (v_{k1}, \dots, v_{kL})'$ and singular values $\varrho_k \neq 0$, $k = 1, \dots, L$. Put $\tilde{\mathbf{u}}_k = (u_{k1}, \dots, u_{kL})'$. Then

$$(\mathbf{C}_{nj}|\mathbf{0}_{d_n \times (d_n - L)})\mathbf{G} = \sum_{k=1}^{L} \varrho_k \begin{pmatrix} \mathbf{c}_{nj}^{(1)\prime} \tilde{\mathbf{u}}_k \\ \vdots \\ \mathbf{c}_{nj}^{(d_n)\prime} \tilde{\mathbf{u}}_k \end{pmatrix} \mathbf{v}_k' = \tilde{\mathbf{C}}_{nj},$$

where $\tilde{\mathbf{C}}_{nj} = (\tilde{c}_{nj}^{(i,l)})_{i,l}$ is the $d_n \times L$ matrix with entries

$$\tilde{c}_{nj}^{(i,l)} = \sum_{k=1}^{L} \varrho_k \mathbf{c}_{nj}^{(i)\prime} \tilde{\mathbf{u}}_k v_{kl}$$

in row $i=1,\ldots,d_n$ and column $l=1,\ldots,L$. Since L is finite, $\|\tilde{\mathbf{u}}_k\|_{\ell_1} \ll 1$ and $c_{nj}^{(\nu,\ell)} \ll (j\vee 1)^{-3/2-\vartheta}$ by Assumption 3, it follows that the coefficients $\tilde{c}_{nj}^{(i,l)}$ satisfy 3 as well. This verifies (i). To see assertion (ii), observe that

$$\mathbf{C}_{nj}\boldsymbol{\varepsilon}_{i-j} = (\mathbf{C}_{nj}|\mathbf{0}_{d_n\times(d_n-L)})\begin{pmatrix}\mathbf{S}_L\\ \mathbf{0}_{(d_n-L)\times L}\end{pmatrix}$$

and let $\mathbf{G}$ be the right matrix factor. If $\mathbf{S}_L=\sum_{k=1}^L \varrho_k\mathbf{u}_k\mathbf{v}_k'$ is the singular value decomposition of $\mathbf{S}_L$ with uniformly ℓ_1-bounded left-singular vectors, then $\sum_{k=1}^L \varrho_k\begin{pmatrix}\mathbf{u}_k\\ \mathbf{0}_{d_n-L}\end{pmatrix}\mathbf{v}_k' = \begin{pmatrix}\mathbf{S}_L\\ \mathbf{0}_{(d_n-L)\times L}\end{pmatrix}$ yields the decomposition of $\mathbf{G}$ with uniformly ℓ_1-bounded left-singular vectors. ■

As a preparation to define approximating martingales for D_{nk}, we need to introduce the following quantities. For $j,k=0,1,2,\ldots$ let

$$f_{k,j}^{(l_1,l_2)}(\mathbf{v}_n,\mathbf{w}_n) = z_{j,\mathbf{v}_n}^{(l_1)}z_{j+k,\mathbf{w}_n}^{(l_2)} + z_{j,\mathbf{w}_n}^{(l_1)}z_{j+k,\mathbf{v}_n}^{(l_2)}\mathbb{1}_{k\neq 0,l_1=l_2},\quad \forall\, l_1,l_2=1,\ldots,L,$$

and, for $i=0,1,2,\ldots$,

$$\tilde{f}_{k,i}^{(\cdot)}(\mathbf{v}_n,\mathbf{w}_n) := \sum_{j=i}^{\infty} f_{k,j}^{(\cdot)}(\mathbf{v}_n,\mathbf{w}_n).$$

For brevity of notation, we ignore the dependence of $f_{k,j}^{(\cdot)}(\mathbf{v}_n,\mathbf{w}_n)$ and $\tilde{f}_{k,j}^{(\cdot)}(\mathbf{v}_n,\mathbf{w}_n)$ on $\mathbf{v}_n$ and $\mathbf{w}_n$ if has no relevance. Later in the proofs we shall need the following preparatory estimates for the above quantities. We show that they are controlled under Assumption 1 (i) by virtue of the following Lemma 2 which is proved in Bours (2019).

Lemma 2. *Suppose that $\mathbf{v}_n$ and $\mathbf{w}_n$ have uniformly bounded ℓ_1-norm in the sense of Equation (1). Then Assumption 1 (i) implies for all $l_1,l_2=1,\ldots,L$*

$$\sup_{n\in\mathbb{N}}\sum_{i=1}^{\infty}\sum_{k=0}^{\infty}(\tilde{f}_{k,i}^{(\cdot)}-\tilde{f}_{k,i+n'}^{(\cdot)})^2 \le C(n')^{1-\vartheta},\quad \forall\, n'=1,2,\ldots, \tag{A1}$$

$$\sup_{n\in\mathbb{N}}\sum_{k=1}^{n'}\sum_{r=0}^{\infty}(\tilde{f}_{r+k,0}^{(\cdot)})^2 \le C(n')^{1-\vartheta},\quad \forall\, n'=1,2,\ldots, \tag{A2}$$

$$\sup_{n\in\mathbb{N}}\sum_{j=1}^{n'}\sum_{k=0}^{\infty}(\tilde{f}_{k,j}^{(\cdot)})^2 \le C(n')^{1-\vartheta},\quad \forall\, n'=1,2,\ldots, \tag{A3}$$

where the constant C may differ from line to line but does not depend on n. Furthermore, the constants depend on the weighting vectors only through their ℓ_1-norms. There exist some

$$\alpha_n^2=\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n)\ge 0,\quad \text{for } n\ge 1, \tag{A4}$$

such that for all $n',m'=0,1,\ldots$ we have

$$\sum_{l_1,l_2=1}^{L}\sum_{j=1}^{n'}\sum_{k=0}^{j-1}\left(\tilde{f}_{k,0}^{(l_1,l_2)}\right)^2\left(\mathrm{E}\left[\left(\varepsilon_{m'+j}^{(l_1)}\right)^2\left(\varepsilon_{m'+j-k}^{(l_2)}\right)^2\right]-\mathrm{E}\left[\varepsilon_{m'+j}^{(l_1)}\varepsilon_{m'+j}^{(l_2)}\right]^2\mathbb{1}_{\{k=0\}}\right)-n'\,\alpha_n^2$$

$$= \sum_{l_1,l_2=1}^{L} \sum_{j=1}^{n'} \sum_{\bar{k}=1}^{j} \left(\tilde{f}_{j-\bar{k},0}^{(l_1,l_2)}\right)^2 \left(\mathrm{E}\left[\left(\varepsilon_{m'+j}^{(l_1)}\right)^2 \left(\varepsilon_{m'+\bar{k}}^{(l_2)}\right)^2 \right] - \mathrm{E}\left[\varepsilon_{m'+j}^{(l_1)}\varepsilon_{m'+j}^{(l_2)}\right]^2 \mathbb{1}_{\{\bar{k}=j\}} \right) - n'\ \alpha_n^2$$
$$\leq C(n')^{1-\vartheta}. \tag{A5}$$

Furthermore, if $\mathbf{v}_n, \mathbf{w}_n, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n$, *for* $n \geq 1$, *have uniformly bounded* ℓ_1*-norms, then there exists*

$$\beta_n^2 = \beta_n(\mathbf{v}_n, \mathbf{w}_n, \tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n), \quad \text{for } n \geq 1, \tag{A6}$$

with

$$\sum_{l_1,l_2=1}^{L} \sum_{j=1}^{n'} \sum_{k=0}^{j-1} \tilde{f}_{k,0}^{(l_1,l_2)}(\mathbf{v}_n, \mathbf{w}_n)\tilde{f}_{k,0}^{(l_1,l_2)}(\tilde{\mathbf{v}}_n, \tilde{\mathbf{w}}_n)\Bigg(\mathrm{E}\left[\left(\varepsilon_{m'+j}^{(l_1)}\right)^2 \left(\varepsilon_{m'+j-k}^{(l_2)}\right)^2 \right]$$
$$- \mathrm{E}\left[\varepsilon_{m'+j}^{(l_1)}\varepsilon_{m'+j}^{(l_2)}\right]^2 \mathbb{1}_{\{k=0\}} \Bigg) - n'\ \beta_n^2 \overset{n',m'}{\ll} C(n')^{1-\vartheta}. \tag{A7}$$

A.2 Martingale approximation

Let us first examine the bilinear form $D_{nk} = \mathbf{v}_n'(\boldsymbol{S}_{nk} - \mathrm{E}(\boldsymbol{S}_{nk}))\mathbf{w}_n$. Notice that we have

$$\boldsymbol{S}_{nk} \overset{(3)}{=} \sum_{i=1}^{k} \sum_{l_1,l_2=1}^{L} \sum_{j_1,j_2=0}^{\infty} \left(c_{nj_1}^{(\nu,l_1)} c_{nj_2}^{(\mu,l_2)} \right)_{1\leq\nu,\mu\leq d_n} \varepsilon_{i-j_1}^{(l_1)} \varepsilon_{i-j_2}^{(l_2)}$$

and

$$\mathrm{E}(\boldsymbol{S}_{nk}) = \sum_{i=1}^{k} \sum_{l_1,l_2=1}^{L} \sum_{j_1,j_2=0}^{\infty} \left(c_{nj_1}^{(\nu,l_1)} c_{nj_2}^{(\mu,l_2)} \right)_{1\leq\nu,\mu\leq d_n} \mathrm{E}\left[\varepsilon_{i-j_1}^{(l_1)} \varepsilon_{i-j_2}^{(l_2)} \right].$$

Introduce L linear processes $Z_i^l(\mathbf{w}_n)$, $l = 1, \ldots, L$, associated to a weighting vector $\mathbf{w}_n$ by

$$Z_i^l(\mathbf{w}_n) := \sum_{j=0}^{\infty} z_{j,\mathbf{w}_n}^{(l)} \varepsilon_{i-j}^{(l)}, \quad \text{with} \quad z_{j,\mathbf{w}_n}^{(l)} := \sum_{p=1}^{d_n} w_p\ c_{nj}^{(p,l)}. \tag{A8}$$

Recall that $Y_i^{(p)} = \sum_{l=1}^{L} \sum_{j=0}^{\infty} c_{nj}^{(p,l)} \varepsilon_{i-j}^{(l)}$, $p = 1, \ldots, d_n$, such that

$$\mathbf{w}_n'\mathbf{Y}_{ni} = \sum_{p=1}^{d_n} w_p \sum_{l=1}^{L} \sum_{j=0}^{\infty} c_{nj}^{(p,l)} \varepsilon_{i-j}^{(l)} = \sum_{l=1}^{L} \sum_{j=0}^{\infty} z_{j,\mathbf{w}_n}^{(l)} \varepsilon_{i-j}^{(l)} = \sum_{l=0}^{L} Z_i^l(\mathbf{w}_n).$$

Consequently,

$$\mathbf{v}_n'\boldsymbol{S}_{nk}\mathbf{w}_n = \sum_{i=1}^{k} \mathbf{v}_n'\mathbf{Y}_{ni}\mathbf{w}_n'\mathbf{Y}_{ni} = \sum_{i=1}^{k} \sum_{l_1=1}^{L} \sum_{l_2=1}^{L} Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n),$$

leading to the representation

$$D_{nk} = \mathbf{v}_n'\boldsymbol{S}_{nk}\mathbf{w}_n - \mathrm{E}(\mathbf{v}_n'\boldsymbol{S}_{nk}\mathbf{w}_n)$$
$$= \sum_{i=1}^{k} \sum_{l_1,l_2=1}^{L} \left(Z_i^{l_1}(\mathbf{v}_n)\ Z_i^{l_2}(\mathbf{w}_n) - \mathrm{E}\left[Z_i^{l_1}(\mathbf{v}_n)\ Z_i^{l_2}(\mathbf{w}_n) \right] \right).$$

We are now in the position to define the martingales. Let $\mathcal{F}_m := \sigma(\epsilon_i^{(1)}, \dots, \epsilon_i^{(L)} : i \le m), m \ge 1$, be the natural filtration. It has been shown in Steland and von Sachs (2017) that for $L = 1$ an approximating $\mathcal{F}_m$-martingale for $D_{nm}(\mathbf{v}_n, \mathbf{w}_n)$ is given by

$$M_m^{(n,1)} = M_m^{(n,1)}(\mathbf{v}_n, \mathbf{w}_n) = \sum_{i=0}^{m} \sum_{k=0}^{\infty} \tilde{f}_{k,0}^{(1)} \left[\epsilon_i^{(1)} \epsilon_{i-k}^{(1)} - \mathrm{E}(\epsilon_i^{(1)})^2 \mathbb{1}_{k=0} \right].$$

For arbitrary L we consider

$$\begin{aligned} M_m^{(n)}(\mathbf{v}_n, \mathbf{w}_n) &= \sum_{i=0}^{m} \sum_{l_1,l_2=1}^{L} \sum_{k=0}^{\infty} \tilde{f}_{k,0}^{(l_1,l_2)}(\mathbf{v}_n, \mathbf{w}_n) \left(\epsilon_i^{(l_1)} \epsilon_{i-k}^{(l_2)} - \mathrm{E}\left[\epsilon_i^{(l_1)} \epsilon_i^{(l_2)} \right] \mathbb{1}_{k=0} \right) \\ &= \sum_{i=0}^{m} \sum_{l_1=1}^{L} \sum_{k=0}^{\infty} \left[\tilde{f}_{k,0}^{(l_1,l_1)} \left(\epsilon_i^{(l_1)} \epsilon_{i-k}^{(l_1)} - \mathrm{E}\left[\left(\epsilon_i^{(l_1)} \right)^2 \right] \mathbb{1}_{k=0} \right) + \sum_{\substack{l_2=1 \\ l_2 \neq l_1}}^{L} \tilde{f}_{k,0}^{(l_1,l_2)} \epsilon_i^{(l_1)} \epsilon_{i-k}^{(l_2)} \right] \end{aligned}$$

for $m, n \ge 0$. The corresponding càdlàg processes are given by

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

Since by Assumption 2 any two innovations $\epsilon_i^{(l_1)}$ and $\epsilon_i^{(l_2)}$ are uncorrelated for $l_1 \neq l_2$, it is easy to see that for each fixed $n \in \mathbb{N}$ the random variables $M_m^{(n)}(\mathbf{v}_n, \mathbf{w}_n)$ satisfy the martingale property

$$\begin{aligned} &\mathrm{E}[M_m^{(n)}(\mathbf{v}_n, \mathbf{w}_n) | \mathcal{F}_{m-1}] \\ &= \sum_{i=0}^{m-1} \sum_{l_1,l_2=1}^{L} \sum_{k=0}^{\infty} \tilde{f}_{k,0}^{(l_1,l_2)} \left(\epsilon_i^{(l_1)} \epsilon_{i-k}^{(l_2)} - \mathrm{E}\left[\epsilon_i^{(l_1)} \epsilon_i^{(l_2)} \right] \mathbb{1}_{k=0} \right) \\ &+ \sum_{l_1,l_2=1}^{L} \left(\sum_{k=1}^{\infty} \tilde{f}_{k,0}^{(l_1,l_2)} \underbrace{\mathrm{E}[\epsilon_m^{(l_1)}]}_{=0} \epsilon_{m-k}^{(l_2)} + \tilde{f}_{0,0}^{(l_1,l_2)} \underbrace{(\mathrm{E}[\epsilon_m^{(l_1)} \epsilon_m^{(l_2)}] - \mathrm{E}[\epsilon_m^{(l_1)} \epsilon_m^{(l_2)}])}_{=0} \right) \\ &= M_{m-1}^{(n)}(\mathbf{v}_n, \mathbf{w}_n), \quad \text{for } m \ge 0. \end{aligned}$$

The associated martingale differences are given by

$$\begin{aligned} &M_{n'+m'}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) - M_{m'}^{(n)}(\mathbf{v}_n, \mathbf{w}_n) \\ &= \sum_{i=m'+1}^{m'+n'} \sum_{l_1,l_2=1}^{L} \sum_{k=0}^{\infty} \tilde{f}_{k,0}^{(l_1,l_2)}(\mathbf{v}_n, \mathbf{w}_n) \left(\epsilon_i^{(l_1)} \epsilon_{i-k}^{(l_2)} - \mathrm{E}\left[\epsilon_i^{(l_1)} \epsilon_i^{(l_2)} \right] \mathbb{1}_{k=0} \right), \end{aligned} \tag{A9}$$

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

In what follows, we study the covariance structure of those martingales and establish formulas for them in terms of the moments of the innovations and the coefficients $\tilde{f}_{k,i}^{(\cdot)}$. Let $(\mathbf{v}_n^{(r)}, \mathbf{w}_n^{(r)})$, $r = 1, 2, \dots$, be weighting vectors. Denote $\tilde{f}_{k,i}^{(\cdot)}(r) = \tilde{f}_{k,i}^{(\cdot)}(\mathbf{v}_n^{(r)}, \mathbf{w}_n^{(r)})$ and let $M_k^{(n)}(r)$, $k \ge 0$, be the corresponding martingales, $r = 1, 2, \dots$. A straightforward but lengthy calculation shows that

$$
\begin{aligned}
&\mathrm{Cov}(M^{(n)}_{m'+n'}(r) - M^{(n)}_{m'}(r), M^{(n)}_{m'+n'}(s) - M^{(n)}_{m'}(s)) \\
&= \sum_{j=m'+1}^{m'+n'} \left[\sum_{l_1,l_2=1}^{L} \left(\tilde{f}^{(l_1,l_2)}_{0,0}(r)\tilde{f}^{(l_1,l_2)}_{0,0}(s) \left\{ (\gamma_{l_1} - \sigma^4_{l_1,m'+j'})\mathbf{1}_{l_1=l_1} + \sigma^2_{l_1,m'+j}\sigma^2_{l_2,m'+j}\mathbf{1}_{l_1\neq l_2} \right\} \right.\right. \\
&\quad \left.\left. + \sum_{k=1}^{\infty} \tilde{f}^{(l_1,l_2)}_{k,0}(r)\tilde{f}^{(l_1,l_2)}_{k,0}(s)\sigma^2_{l_1,m'+j}\sigma^2_{l_2,m'+j-k} \right) \right]. \qquad \text{(A10)}
\end{aligned}
$$

The following lemma follows using similar arguments as in Steland (2020).

Lemma 3. *Suppose that $\varepsilon^{(l)}_i = \varepsilon^{(l)}_{ni}$ depend on n and put $\sigma^2_{l,ni} = \mathrm{E}(\varepsilon^{(l)}_{ni})^2$ and $\gamma_{l,ni} = \mathrm{E}(\varepsilon^{(l)}_{ni})^4$ for all i and n. Assume that*

$$
\frac{1}{n'} \sum_{i=1}^{n'} i|\sigma^2_{l,ni} - s^2_l| \overset{n,n'}{\ll} (n')^{-\beta} \qquad \text{(A11)}
$$

and

$$
\frac{1}{n'} \sum_{i=1}^{n'} i|\gamma^2_{l,ni} - \gamma^2_l| \overset{n,n'}{\ll} (n')^{-\beta} \qquad \text{(A12)}
$$

for constants $s^2_l \in (0,\infty)$ and $\gamma_l \in \mathbb{R}$, $l = 1, \dots, L$, for some $1 < \beta < 2$ with $1 + \vartheta < \beta$. Then, for $n, n' \geq 1$ and $m' \geq 0$

$$
|Cov(M^{(n)}_{m'+n'}(r) - M^{(n)}_{m'}(r), M^{(n)}_{m'+n'}(s) - M^{(n)}_{m'}(s)) - n'\beta^2_n(r,s)| \overset{n,n',m'}{\ll} (n')^{1-\vartheta}, \qquad \text{(A13)}
$$

where for weighting vectors $(\mathbf{v}_r, \mathbf{w}_r) = (\mathbf{v}_{nr}, \mathbf{w}_{nr})$ and $(\mathbf{v}_s, \mathbf{w}_s) = (\mathbf{v}_{ns}, \mathbf{w}_{ns})$

$$
\begin{aligned}
\beta^2_n(\mathbf{v}_r, \mathbf{w}_r, \mathbf{v}_s, \mathbf{w}_s) = &\sum_{l_1,l_2=1}^{L} \tilde{f}^{(l_1,l_2)}_{0,0}(r)\tilde{f}^{(l_1,l_2)}_{0,0}(s)\{(\gamma_{l_1} - s^4_{l_1})\mathbf{1}_{l_1=l_2} + s^2_{l_1}s^2_{l_2}\mathbf{1}_{l_1\neq l_2}\} \\
&+ s^2_{l_1}s^2_{l_2}\sum_{k=1}^{\infty} \tilde{f}^{(l_1,l_2)}_{k,0}(r)\tilde{f}^{(l_1,l_2)}_{k,0}(s), \qquad \text{(A14)}
\end{aligned}
$$

for $l = 1, \dots, L$.

Lemma 4. *We have*

$$
\begin{aligned}
E_{n,n'} &:= \left\| \mathrm{E}\left[\left(M^{(n)}_{m'+n'}(r) - M^{(n)}_{m'}(r) \right) \left(M^{(n)}_{m'+n'}(s) - M^{(n)}_{m'}(s) \right) \mid \mathcal{F}_m \right] - n'\beta^2_n(r,s) \right\|_{L_1} \\
&\overset{n,m'}{\ll} (n')^{1-\vartheta/2}.
\end{aligned}
$$

Let us now study how well the martingale differences approximate the bilinear form. Associated to the bilinear form $D_{nk}(\mathbf{v}_n, \mathbf{w}_n)$ we define

$$
D^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n) := \sum_{i=m'+1}^{m'+n'} \sum_{l_1,l_2=1}^{L} Z^{l_1}_i(\mathbf{v}_n)Z^{l_2}_i(\mathbf{w}_n) - \mathrm{E}\left[Z^{l_1}_i(\mathbf{v}_n)Z^{l_2}_i(\mathbf{w}_n) \right], \qquad \text{(A15)}
$$

for $m', n' \geq 0$ and consider the decomposition

$$D^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n) = M^{(n)}_{n'+m'}(\mathbf{v}_n, \mathbf{w}_n) - M^{(n)}_{m'}(\mathbf{v}_n, \mathbf{w}_n) + R^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n). \tag{A16}$$

In order to justify the approximation of $D^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n)$ by the lag m' martingale differences, we need to show that the approximation error $R^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n)$ is asymptotically negligible, as $n', m' \to \infty$. The following lemma shows that this holds under the conditions of Theorem 2 in the L_2-sense and also provides a convergence rate.

Lemma 5. *Under the conditions of Theorem 2 we have*

$$\mathrm{E}\left[R^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n)^2\right] \overset{m',n'}{\ll} (n')^{1-\vartheta}, \tag{A17}$$

for all $m', n' = 0, 1, 2, \ldots$, where $0 < \vartheta < \frac{1}{2}$ is the constant defined in Assumption 3.

Clearly, Lemma 5 ensures that, by an application of the law of total expectation,

$$\left\|\mathrm{E}\left[R^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n)^2 | \mathcal{F}_{m'}\right]\right\|_1 = \mathrm{E}\left[R^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n)^2\right] \overset{n',m'}{\ll} (n')^{1-\vartheta}. \tag{A18}$$

A.3 Proof of Theorem 2

In the following we show that the bilinear forms $D_{nk}(\mathbf{v}_n, \mathbf{w}_n)$ satisfy a strong approximation result with error term $O\left(t^{\frac{1}{2}-\kappa}\right)$, for some $\kappa > 0$, by relying on the following result of Philipp (1986), see also Eberlein (1986).

Theorem 7. *Let $\xi^{(n)}_i, i = 1, \ldots, n$, be an array of random variables with values in $\mathbb{R}^d$ where $d \leq \infty$. Furthermore let $\{\mathcal{F}^{(n)}_i, i \geq 1\}$ be a nondecreasing sequence of σ-fields such that $\xi^{(n)}_i$ is $\mathcal{F}^{(n)}_i$-measurable. Suppose that*

$$D^{(n)}_{n',m'} := \sum_{i=m'+1}^{m'+n'} \xi^{(n)}_i,$$

satisfies for $m' \geq 0, n' \geq 1$ the following conditions:

(i) $\mathrm{E}|\mathrm{E}[D^{(n)}_{n',m'} | \mathcal{F}^{(n)}_{m'}]| \overset{n',m'}{\ll} (n')^{\frac{1}{2}-\varepsilon}$, *a.s., for some $\varepsilon > 0$.*
(ii) $\sup_{i \geq 0} \mathrm{E}|\xi^{(n)}_i|^{4+\delta} < \infty$, *for some $\delta > 0$.*
(iii) There exists a variance parameter $\alpha^2_n \geq 0$ such that

$$\mathrm{E}\left|\mathrm{E}\left[(D^{(n)}_{n',m'})^2 | \mathcal{F}^{(n)}_{m'}\right] - n'\alpha^2_n\right| \overset{n',m'}{\ll} (n')^{1-\varepsilon}, \text{ a.s., for some } \varepsilon > 0.$$

Then, without changing its distribution, we can redefine the sequence $\{\xi^{(n)}_i : i \geq 1\}$ on some new probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ on which there exists a process $\{\tilde{D}^{(n)}_{n'} : n' \geq 0\}$ and a standard Brownian motion $\{\tilde{B}^{(n)}_t : t \geq 0\}$ with incremental variance α^2_n, such that

$$\{\tilde{D}^{(n)}_{n'} : n' \geq 0\} \overset{d}{=} \{D^{(n)}_{n',0} : n' \geq 0\}$$

and for some $\lambda > 0$

$$\left|\tilde{D}^{(n)}_{\lfloor t\rfloor} - \alpha_n \tilde{B}^{(n)}_t\right| \overset{t}{\ll} t^{\frac{1}{2}-\lambda}, \quad \text{for all } t > 0 \quad \tilde{P}\text{-a.s.}$$

Let us now show that the array of random variables

$$\xi_i^{(n)} = \xi_i^{(n)}(\mathbf{v}_n, \mathbf{w}_n) = \sum_{l_1,l_2=1}^{L} Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n) - \mathrm{E}\left[Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n)\right], \tag{A19}$$

$1 \le i \le n$, $n \ge 1$, satisfies the conditions (i)–(iii) of Theorem 7.

Condition (i): Condition (i) is readily verified through Lemma 5 and the martingale property because for some $\frac{\vartheta}{2} > 0$ we have

$$\begin{aligned}
&\left\|\mathrm{E}\left[D^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n)|\mathcal{F}_{m'}\right]\right\|_1 \\
&\overset{(43)}{=} \left\|M^{(n)}_{m'}(\mathbf{v}_n, \mathbf{w}_n) - M^{(n)}_{m'}(\mathbf{v}_n, \mathbf{w}_n) + \mathrm{E}\left[R^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n)|\mathcal{F}_{m'}\right]\right\|_1 \\
&= \left\|\mathrm{E}\left[R^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n)|\mathcal{F}_{m'}\right]\right\|_1 \le \sqrt{\left\|\mathrm{E}\left[R^{(n)}_{n',m'}(\mathbf{v}_n, \mathbf{w}_n)^2|\mathcal{F}_{m'}\right]\right\|_1} \\
&\underset{(45)}{\overset{n',m'}{\ll}} (n')^{\frac{1}{2}-\frac{\vartheta}{2}}, \ a.s.
\end{aligned}$$

Condition (ii): Using the c_r-Inequality $\mathrm{E}|X+Y|^r \le c_r(\mathrm{E}|X|^r + \mathrm{E}|Y|^r)$, where $c_r = 2^{r-1}$, for $r > 1$, we get

$$\begin{aligned}
\mathrm{E}\left[|\xi_i^{(n)}|^{2+\delta}\right] &= \mathrm{E}\left[\left|\sum_{l_1,l_2=1}^{L} Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n) - \mathrm{E}\left[Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n)\right]\right|^{2+\delta}\right] \\
&\le 2^{(1+\delta)\frac{L}{2}} \sum_{l_1,l_2=1}^{L} \mathrm{E}\left[\left|Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n) - \mathrm{E}\left[Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n)\right]\right|^{2+\delta}\right].
\end{aligned}$$

Therefore, the assertion follows immediately, if we prove that

$$\sup_{i \ge 0} \mathrm{E}\left[\left|Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n) - \mathrm{E}\left[Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n)\right]\right|^{2+\delta}\right] < \infty,$$

for all $l_1, l_2 = 1, \dots, L$. Let us first mention that with the assumption on our innovations the cross-sectional correlations of $Z_i^{(l)}(\cdot)$ are finite since we have

$$\mathrm{E}\left[Z_i^{l_1}(\cdot)\right] = \sum_{j=0}^{\infty} z_{j,\cdot}^{(m)} \mathrm{E}\left[\varepsilon_{i-j}^{(m)}\right] = 0$$

and for $l_1 = l_2$ we get

$$\mathrm{E}\left[Z_i^l(\mathbf{v}_n) Z_i^l(\mathbf{w}_n)\right] = \sum_{j=0}^{\infty} z_{j,\mathbf{v}_n}^{(l)} z_{j,\mathbf{w}_n}^{(l)} \underbrace{\mathrm{E}\left[(\varepsilon_{i-j}^{(l)})^2\right]}_{\le C < \infty} \overset{(48)}{\le} C \sum_{j=0}^{\infty} j^{-\frac{3}{2}-\vartheta} \le C, \tag{A20}$$

where $C<\infty$ is a suitable constant. Here, and in what follows, we use the fact that for $j \geq 0$ and $l=1, \dots, L$,

$$\sup_{n\in\mathbb{N}} \left| z_{j,\mathbf{w}_n}^{(l)} \right|^2 = \sup_{n\in\mathbb{N}} \left| \sum_{p=1}^{d_n} w_p \; c_{nj}^{(p,l)} \right|^2 \leq \sup_{n\in\mathbb{N}} \left(\sum_{p=1}^{d_n} |w_p|^2 \right) \left| \max_{1\leq p\leq d_n} c_{nj}^{(p,l)} \right|^2$$
$$\overset{\text{Ann}(C)}{\leq} C \sup_{n\in\mathbb{N}} \|\mathbf{w}_n\|_{\ell_1}^2 (j \vee 1)^{-\frac{3}{2}-\vartheta} \leq C_w \; (j \vee 1)^{-\frac{3}{2}-\vartheta}. \tag{A21}$$

Next, we obtain for $l_1 \neq l_2$

$$\mathrm{E}\left[Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n)\right] = \sum_{j=0}^{\infty} z_{j,\mathbf{v}_n}^{(l_1)} \; z_{j,\mathbf{w}_n}^{(l_2)} \; \underbrace{\mathrm{E}\left[\varepsilon_{i-j}^{(l_1)}\right] \mathrm{E}\left[\varepsilon_{i-j}^{(l_2)}\right]}_{=0} = 0. \tag{A22}$$

Now we use the c_r-inequality and the Cauchy–Schwarz inequality and obtain

$$\sup_{i\geq 0} \mathrm{E}\left[\left| Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n) - \mathrm{E}\left[Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n)\right] \right|^{2+\delta} \right]$$
$$\underset{(49)}{\overset{(47)}{\leq}} \sup_{i\geq 0} 2^{1+\delta} \; \mathrm{E}\left(\left[|Z_i^{l_1}(\mathbf{v}_n) Z_i^{l_2}(\mathbf{w}_n)|^{2+\delta} \right] + \mathrm{E}\left[C^{2+\delta}\right] \right)$$
$$\overset{CS}{\leq} \sup_{i\geq 0} 2^{1+\delta} \sqrt{\mathrm{E}\left[|Z_i^{l_1}(\mathbf{v}_n)|^{4+2\delta} \right]} \sqrt{\mathrm{E}\left[|Z_i^{l_2}(\mathbf{w}_n)|^{4+2\delta} \right]} + C.$$

Therefore, Condition (ii) is immediately verified if we show that

$$\sup_{i\geq 0} \mathrm{E}\left[|Z_i^l(\cdot)|^{4+2\delta} \right] < \infty, \quad \forall \, l = 1, \dots, L.$$

This is easy to show if we use the inequality by Marcinkiewicz and Zygmund (1937): If $X_i, i=1, \dots, n$ are independent random variables with $\mathrm{E}(X_i) = 0$, then for every $p \geq 1$ there exist positive constants A_p and B_p depending only upon p for which

$$A_p \; \mathrm{E}\left[\left(\sum_{i=1}^n |X_i|^2 \right)^{p/2} \right] \leq \mathrm{E}\left[\left| \sum_{i=1}^n X_i \right|^p \right] \leq B_p \; \mathrm{E}\left[\left(\sum_{i=1}^n |X_i|^2 \right)^{p/2} \right].$$

This inequality also holds for $n=\infty$ if both $\sum_{i=1}^n X_i$ and $\sum_{i=1}^n X_i^2$ converge a.s., as $n \to \infty$ (see Chow and Teicher, 1997, theorem 2 and corollary 3, pp. 386–387).

We can now follow the arguments of Kouritzin (1995, pp. 351-352). Let $\xi = \frac{\delta}{2}$. Without loss of generality we assumed $\delta < 2$ because in this case we have $p := \frac{1+\xi}{2} < 1$ and the function $f(x)=x^p$ is concave, for $x \geq 0$. Then Jensen's inequality yields $\mathrm{E}[X^{1+\xi}] = \mathrm{E}[(X^2)^{\frac{1+\xi}{2}}] \leq (\mathrm{E}[X^2])^{\frac{1+\xi}{2}}$. For simplicity we ignore the dependence of $Z_k(\mathbf{v}_n), z_{j,\cdot}^{(l)}$ and $\varepsilon_{i-j}^{(l)}$ on $\mathbf{v}_n$ and l. Using the Marcinkiewicz–Zygmund inequality (MZ) and Jensen's inequality (J), we find

$$\mathrm{E}[|Z_i^l(\cdot)|^{4+2\delta}] = \mathrm{E}\left| \sum_{j=0}^{\infty} z_j \varepsilon_{i-j} \right|^{4+\delta} \underset{(\mathrm{MZ})}{\overset{k}{\ll}} \mathrm{E}\left[\left| \sum_{j=0}^{\infty} z_j^2 \varepsilon_{i-j}^2 \right|^{2+\xi} \right]$$

$$= \mathrm{E}\left[\left(\sum_{j=0}^{\infty} z_j^2 \varepsilon_{i-j}^2\right)\left|\sum_{j=0}^{\infty} z_j^2 \varepsilon_{i-j}^2\right|^{1+\xi}\right]$$

$$\overset{k}{\ll} \sum_{j=0}^{\infty} \mathrm{E}\left[z_j^2 \varepsilon_{i-j}^2 \left(z_j^2 \varepsilon_{i-j}^2\right)^{1+\xi}\right] + \sum_{j=0}^{\infty} \mathrm{E}\left[z_j^2 \varepsilon_{i-j}^2 \left|\sum_{n \neq j} z_n^2 \varepsilon_{i-n}^2\right|^{1+\xi}\right]$$

and therefore

$$\mathrm{E}\left[|Z_i^l(\cdot)|^{4+2\delta}\right] = \sum_{j=0}^{\infty} \mathrm{E}\left[|z_j^2 \varepsilon_{i-j}^2|^{2+\xi}\right] + \sum_{j=0}^{\infty}\left(z_j^2 \mathrm{E}\left[\varepsilon_{i-j}^2\right]\ \mathrm{E}\left[\left|\sum_{n \neq j} z_n^2 \varepsilon_{i-n}^2\right|^{1+\xi}\right]\right)$$

$$\overset{k}{\ll} \sum_{j=0}^{\infty} \mathrm{E}\left[|z_j^2 \varepsilon_{i-j}^2|^{2+\xi}\right] + \sum_{j=0}^{\infty} z_j^2\ \mathrm{E}\left[\varepsilon_{i-j}^2\right]\ \mathrm{E}\left[\left|\sum_{n=0}^{\infty} z_n^2 \varepsilon_{i-n}^2\right|^{1+\xi}\right]$$

$$\underset{(J)}{\overset{k}{\ll}} \sum_{j=0}^{\infty} (z_j^2)^{2+\xi} \underbrace{\mathrm{E}\left[|\varepsilon_{i-j}^2|^{2+\xi}\right]}_{\overset{(4)}{\leq} C<\infty} + \sum_{j=0}^{\infty} z_j^2 \underbrace{\mathrm{E}[\varepsilon_{i-j}^2]}_{\overset{(4)}{\leq} C<\infty} \left|\mathrm{E}\left|\sum_{n=0}^{\infty} z_n^2 \varepsilon_{i-n}^2\right|^2\right|^{\frac{1+\xi}{2}}$$

$$\underset{(48)}{\overset{k}{\ll}} \underbrace{\sum_{j=0}^{\infty} (j \vee 1)^{-\left(\frac{3}{2}+\vartheta\right)(2+\xi)}}_{<\infty,\ \text{since}\ \left(\frac{3}{2}+\vartheta\right)(2+\xi)>1} + \underbrace{\sum_{j=0}^{\infty} (j \vee 1)^{-\frac{3}{2}-\vartheta}}_{<\infty,\ \text{since}\ \frac{3}{2}+\vartheta>1} \left(\mathrm{E}\left|\sum_{n=0}^{\infty} z_n^2 \varepsilon_{k-n}^2\right|^2\right)^{\frac{1+\xi}{2}}$$

$$\overset{k}{\ll} 1.$$

The last inequality holds true since we have

$$\mathrm{E}\left|\sum_{n=0}^{\infty} z_n^2 \varepsilon_{k-n}^2\right|^2 = \sum_{n=0}^{\infty} z_n^4 \mathrm{E}[\varepsilon_{k-n}^4] + \sum_{\substack{n_1,n_2=0\\ n1\neq n2}}^{\infty} z_{n_1}^2 z_{n_2}^2 \mathrm{E}[\varepsilon_{k-n_1}^2]\mathrm{E}[\varepsilon_{k-n_2}^2]$$

$$\leq C\left(\sum_{n=0}^{\infty} z_n^4 + \sum_{\substack{n_1,n_2=0\\ n1\neq n2}}^{\infty} z_{n_1}^2 z_{n_2}^2\right) = C\left(\sum_{n=0}^{\infty} z_n^2\right)^2$$

$$\overset{(48)}{\leq} C\left(\sum_{n=0}^{\infty} (n \vee 1)^{-\frac{3}{2}-\vartheta}\right)^2 < \infty.$$

This finishes the proof and it is shown that $\xi_k^{(n)}$ satisfies Condition (ii).

Condition (iii): For simplicity and clarity of presentation, we ignore the dependence on $(\mathbf{v}_n, \mathbf{w}_n)$ in what follows. Using the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$ (BE), we have

$$\left\|\mathrm{E}\left[(D_{n',m'}^{(n)})^2|\mathcal{F}_{m'}\right]-n'\alpha_n^2\right\|_1 \overset{(43)}{=} \left\|\mathrm{E}\left[\left(M_{n'+m'}^{(n)}-M_{m'}^{(n)}+R_{n',m'}^{(n)}\right)^2|\mathcal{F}_{m'}\right]-n'\alpha_n^2\right\|_1$$
$$\overset{(BE)}{\le} 2\left\|\mathrm{E}\left[\left(M_{n'+m'}^{(n)}-M_{m'}^{(n)}\right)^2|\mathcal{F}_{m'}\right]-n'\alpha_n^2\right\|_1+2\left\|\mathrm{E}\left[\left(R_{n',m'}^{(n)}\right)^2|\mathcal{F}_{m'}\right]\right\|_1.$$

Here the first term is $\overset{n,n',m'}{\ll} (n')^{1-\vartheta/2}$ by Lemma 4 and the second one is $\overset{n,n',m'}{\ll} (n')^{1-\vartheta/2}$ by (A18), see Lemma 5, which shows (iii).

Consequently, Theorem 7 yields the existence of a standard Brownian motion, $B_n(t), t\in[o,\infty)$, such that for some constant C_n and a universal constant $\lambda>0$ we have the approximation

$$|D_{nt}-\alpha_n B_n(t)|\le C_n t^{\frac{1}{2}-\lambda}, \quad \text{for all } t>0, \quad a.s.$$

This bound is sharp enough to yield the strong approximation (17). Denoting the standard Brownian motion on $[0,1]$ associated to B_n, $t\mapsto n^{-\frac{1}{2}}B_n(tn), t\in[0,1]$, again by B_n, we obtain

$$\sup_{t\in[0,1]}\left|n^{-\frac{1}{2}}D_{n,\lfloor nt\rfloor}-\alpha_n B_n\left(\frac{\lfloor nt\rfloor}{n}\right)\right|\le C_n n^{-\lambda}$$

which establishes the statement, provided $C_n n^{-\lambda}=o(1)$. Assertion (17) on the process $\mathcal{D}_n(t)$ follows now easily as well as the result for $\mathcal{D}_n^0(t)$.

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

$$\left\|\mathrm{E}\left[\left(M_{m'+n'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)-M_{m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)\right)^2|\mathcal{F}_{m'}\right]-n'\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n)\right\|_1 \overset{n',m'}{\ll} (n')^{1-\frac{\vartheta}{2}}$$

and

$$\left\|\mathrm{E}\left[D_{n',m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)^2|\mathcal{F}_{m'}\right]-n'\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n)\right\|_1 \overset{n',m'}{\ll} (n')^{1-\frac{\vartheta}{2}}$$

as well as

$$\left\|\mathrm{E}\left[D_{n',m'}^{(n)}(\mathbf{v}_n,\mathbf{w}_n)^2-n'\alpha_n^2(\mathbf{v}_n,\mathbf{w}_n)\right]\right\|_1 \overset{n',m'}{\ll} (n')^{1-\frac{\vartheta}{2}}.$$

The constants appearing above depend on the weighting vectors only through their ℓ_1-norms.

It remains to provide the arguments for the standardized statistics under Assumption 1 (ii) on the projection vectors. First observe that for any sequence $\mathbf{v}_n$ of weights satisfying Assumption 1 (ii) the scaled vectors $d_n^{-1/2}\mathbf{v}_n$ satisfy Assumption 1 (i), since by virtue of Jensen's inequality

$$d_n^{-1/2}\|\mathbf{v}_n\|_{\ell_1}=\sqrt{d_n}\frac{1}{d_n}\sum_{\nu=1}^{d_n}\sqrt{v_{n\nu}^2}\le\sqrt{d_n}\left(d_n^{-1}\sum_{\nu=1}^{d_n}v_{n\nu}^2\right)^{1/2}=\|\mathbf{v}_n\|_{\ell_2}.$$

Since the statistics standardized by $\alpha_n(\mathbf{v}_n,\mathbf{w}_n)$ are invariant under scaling of the weighting vectors, we may therefore apply the above results with $d_n^{-1/2}\mathbf{v}_n$ and $d_n^{-1/2}\mathbf{w}_n$. When using a homogenous and weakly consistent estimator, we may assume that the estimator is strongly consistent when redefined on the new probability space. The difference between $\mathcal{D}_n(t)/\hat{\alpha}_n$ and $\mathcal{D}_n(t)/\alpha_n$ can be treated similarly as given in detail in the proof of Theorem 3, using the fact that $\sup_{0<t<1}|B_n(t)|<\infty$, a.s., $\alpha_n/\hat{\alpha}_n\to 1$, $n\to\infty$, a.s., and $|\hat{\alpha}_n-\alpha_n|/\alpha_n\to 0$, $n\to\infty$, a.s.. Putting things together completes the proof.

A.4 Proof of Theorem 3

Since for $1 \leq i < j \leq n$

$$\sum_{r=1}^{i} D_{nr} - \frac{j}{n}\sum_{\ell=1}^{n} D_{n\ell} - \left(\sum_{r=1}^{i} D_{nr} - \frac{i}{n}\sum_{\ell=1}^{n} D_{n\ell}\right) = \sum_{i<r\leq j} D_{nr} - \frac{j-i}{n}\sum_{\ell=1}^{n} D_{n\ell},$$

we have the representation

$$T_{n,md} = \max_{1\leq i<j\leq n} |\mathcal{D}_n^0(j/n) - \mathcal{D}_n^0(i/n)| = \sup_{0<s<t<1} |\mathcal{D}_n^0(t) - \mathcal{D}_n^0(s)|.$$

Furthermore,

$$\begin{aligned} R_n &= \sup_{0<s<t<1} \alpha_n^{-1}|\mathcal{D}_n^0(t) - \mathcal{D}_n^0(s)| - \sup_{0<s<t<1} |B_n^0(t) - B_n^0(s)| \\ &= \alpha_n^{-1} \sup_{0<s<t<1} (|\mathcal{D}_n^0(t) - \mathcal{D}_n^0(s)| - \sup_{0<s'<t'<1} |B_n^0(t') - B_n^0(s')|) \\ &\leq \alpha_n^{-1} \sup_{0<s<t<1} (|\mathcal{D}_n^0(t) - \mathcal{D}_n^0(s)| - |B_n^0(t) - B_n^0(s)|) \\ &\leq \alpha_n^{-1} \sup_{0<s<t<1} |\mathcal{D}_n^0(t) - \mathcal{D}_n^0(s) - [B_n^0(t) - B_n^0(s)]| \\ &\leq 2\alpha_n^{-1} \sup_{0<s<t<1} |\mathcal{D}_n^0(t) - B_n^0(t)| = o(1), \end{aligned}$$

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

$$\begin{aligned} &\hat{\alpha}_n^{-1} \sup_{0<s<t<1} |\mathcal{D}_n^0(t) - \mathcal{D}_n^0(s)| - \alpha_n^{-1} \sup_{0<s<t<1} |\mathcal{D}_n^0(t) - \mathcal{D}_n^0(s)| \\ &= \frac{\alpha_n - \hat{\alpha}_n}{\hat{\alpha}_n} \alpha_n^{-1} \sup_{0<s<t<1} |\mathcal{D}_n^0(t) - \mathcal{D}_n^0(s)| \\ &= \frac{\alpha_n}{\hat{\alpha}_n}\frac{\alpha_n - \hat{\alpha}_n}{\alpha_n} \alpha_n^{-1} \sup_{0<s<t<1} |B_n^0(t) - B_n^0(s)| + \frac{\alpha_n}{\hat{\alpha}_n}\frac{\alpha_n - \hat{\alpha}_n}{\alpha_n} R_n \\ &= o_P(1), \end{aligned}$$

as $n \to \infty$, in view of the estimate for R_n and since $\hat{\alpha}_n/\alpha_n \stackrel{P}{\to} 1$, α_n is bounded away from 0 for large enough n and $\sup_{0<s<t<1} |B_n^0(t) - B_n^0(s)| = O_P(1)$ by (10). These facts imply

$$T_{n,md} \stackrel{d}{\to} \sup_{0<s<t<1} |B_n^0(t) - B_n^0(s)|, \quad n \to \infty.$$

Now let $p_n \in (0, 1)$ be an arbitrary sequence of probabilities with $p_n \to 0$, $n \to \infty$. Then for large enough n the unique solution u_n of $2^{5/2}\sqrt{\pi}u^3(1 - \Phi(2u) = p_n$ satisfies $u_n \to \infty$, since $u^3(1 - \Phi(2u))$, $u > u_0$, is strictly decreasing for large u_0. We may conclude that

$$\begin{aligned} |\mathrm{P}(T_{n,md} > u_n) - p_n| &= \left|\mathrm{P}(T_{n,md} > u_n) - 2^{5/2}\sqrt{\pi}u_n^3(1 - \Phi(2u_n)\right| \\ &= \left|\mathrm{P}(T_{n,md} > u_n) - \mathrm{P}\left(\sup_{0<s<t<1} |B_n^0(t) - B_n^0(s)| > u_n\right) + o(1)\right| \end{aligned}$$

$$\leq \sup_{z \in \mathbb{R}} \left| \mathrm{P}(T_{n,md} \leq z) - \mathrm{P}\left(\sup_{0<s<t<1} |B_n^0(t) - B_n^0(s)| \leq z \right) \right| + o(1)$$
$$= o(1).$$

A.5 Proof of Theorem 4

We proceed analogously to the proof of Theorem 2 in Section A0.3. We define the multivariate extension of the bilinear form D_{nk} by

$$\mathbf{D}_{nk} = (D_{nk}(j))_{j=1}^{M} = \left(\mathbf{v}_{nj}'(\boldsymbol{S}_{nk} - \mathrm{E}(\boldsymbol{S}_{nk}))\mathbf{w}_{nj} \right)_{j=1}^{M}.$$

Introduce the linear processes $Z_i^l(\mathbf{w}_{nj})$, $l = 1, \ldots, L, j = 1, \ldots, M$, associated to a weighting vector $\mathbf{w}_{nj} = (w_1^{(j)}, \ldots, w_{d_n}^{(j)})'$ by

$$Z_i^l(\mathbf{w}_{nj}) := \sum_{j=0}^{\infty} z_{j,\mathbf{w}_{nj}}^{(l)} \, \epsilon_{i-j}^{(l)}, \quad \text{with} \quad z_{j,\mathbf{w}_{nj}}^{(l)} := \sum_{p=1}^{d_n} w_p^{(j)} \, c_{nj}^{(p,l)}. \tag{A23}$$

For ease of reading we shall, except when proving Theorem 5, omit in our notation the dependence of the random variables Z_i^l and the coefficients $z_{j,\mathbf{w}_{nj}}^{(l)}$ on the sample size. We obtain the representation

$$\mathbf{D}_{nk} = \sum_{i=1}^{k} \sum_{l_1,l_2=1}^{L} \underbrace{\left(Z_i^{l_1}(\mathbf{v}_n) \, Z_i^{l_2}(\mathbf{w}_n) - \mathrm{E}\left[Z_i^{l_1}(\mathbf{v}_n) \, Z_i^{l_2}(\mathbf{w}_n) \right] \right)_{j=1}^{M}}_{=:\xi_i^{(n)}}.$$

Furthermore we put

$$\mathbf{S}_{n',m'}^{(n)} = \sum_{k=m'+1}^{m'+n'} \xi_k^{(n)}, \; n', m' \geq 0.$$

For the Euclidean space $\mathbb{R}^M$ equipped with the usual inner product and the induced vector ℓ_2-norm, the conditions (i) and (iii) are easily checked, similar to the proof of Theorem 2. In order to show condition (ii), we introduce the conditional covariance operator

$$C_{n',m'}^{(n)}(\mathbf{u}) = \mathrm{E}\left[\mathbf{u}' \, \mathbf{S}_{n',m'}^{(n)} \, \mathbf{S}_{n',m'}^{(n)} | \mathcal{F}_{m'} \right], \; \mathbf{u} \in \mathbb{R}^M,$$

and the covariance operator

$$T^{(n)}(\mathbf{u}) = \mathrm{E}\left[\mathbf{u}'\mathbf{B}^{(n)}\mathbf{B}^{(n)} \right] = \sum_{j=1}^{M} u_j (\mathrm{Cov}(B_{n1}, B_{nj}), \ldots, \mathrm{Cov}(B_{nM}, B_{nj}))', \quad \mathbf{u} \in \mathbb{R}^M.$$

We need to check that

$$\mathrm{E} \left\| \frac{1}{n'} C_{n',m'}^{(n)} - C^{(n)} \right\| \overset{n',m'}{\ll} (n')^{-\vartheta}, \quad \text{for some covariance operator } C^{(n)}. \tag{A24}$$

Here the norm is the operator norm $\|Q\| = \sup_{u\in\mathbb{R}^M,\|u\|=1}|\mathbf{u}'Q\mathbf{u}|$ for a linear operator $Q : \mathbb{R}^M \mapsto \mathbb{R}^M$. It follows that (A24) holds true with $C^{(n)} = T^{(n)}$, since we showed in Lemma 2

$$\left\|\mathrm{E}\left[(D^{(n)}_{n',m'}(i))^2|\mathcal{F}_{m'}\right] - n'\alpha_n(\mathbf{v}_{ni},\mathbf{w}_{ni})\right\|_{L_1} \underset{(32)}{\overset{m'}{\ll}} (n')^{1-\vartheta/2}, \quad \forall i = 1, \dots, M,$$

and for all $i,j = 1, \dots, M$ with $i \neq j$

$$\left\|\mathrm{E}\left[D^{(n)}_{n',m'}(i)D^{(n)}_{n',m'}(j)|\mathcal{F}_{m'}\right] - n'\beta_n(\mathbf{v}_{ni},\mathbf{w}_{ni},\mathbf{v}_{nj},\mathbf{w}_{nj})\right\|_{L_1} \underset{(34)}{\overset{m'}{\ll}} (n')^{1-\vartheta/2}. \tag{A25}$$

A.6 Proof of Theorem 5

We have $Y_{nk}(\mathbf{v}) := \mathbf{v_n}'\mathbf{Y}_{nk} \overset{(35)}{=} \sum_{l=0}^{L} Z^l_{nk}(\mathbf{v}_n)$, with $Z^l_{nk}(\mathbf{v}) = \sum_{j=0}^{\infty} z^{(l)}_{nj,\mathbf{v}_n}\epsilon^{(l)}_{k-j}$ and $z^{(l)}_{nj,\mathbf{v}_n} = \sum_{p=1}^{d_n} w_{np}c^{(p,l)}_{nj}$, see (A8), where we explicitly indicate the dependence on n in the notation. Investigating the coefficients $z^{(l)}_{nj,v}$ we obtain, uniformly in n,

$$\left|z^{(l)}_{nj,\mathbf{v}_r}\right| \le \sup_{q\ge1}\left|c^{(q,l)}_{nj}\right|\sum_{p=1}^{d_n}|v_{np}| = \|\mathbf{v}_{nr}\|_{\ell_1}\sup_{q\ge1}\left|c^{(q,l)}_{nj}\right| \ll (j\vee 1)^{-(1+\delta)}$$

and analogously

$$\left|z^{(l)}_{n,j+|h|,\mathbf{v}_r}\right| \le \|\mathbf{v}_{nr}\|_{\ell_1}\sup_{q\ge1}\left|c^{(q,l)}_{n,j+|h|}\right| \ll ((j+|h|)\vee 1)^{-(1+\delta)}.$$

Observe that $Y_{nk}(\mathbf{v})Y_{nk}(\mathbf{w}), k\ge1$, as well as $Y_{nk}(\mathbf{v}_r)Y_k(\mathbf{w}_r)Y_{k+h}(\mathbf{v}_s)Y_{n,k+h}(\mathbf{w}_s), k\ge1$, are strictly stationary for any fixed r, s, and h. The dependence on d will be suppressed in notation. Similar to Steland and von Sachs (2017) we can conclude from the proof of Theorem 2 and Kouritzin (1995, p. 351), that α^2 can be represented as

$$\alpha^2 = \lim_{n\to\infty}\alpha_n^2 = \lim_{n\to\infty}\mathrm{Var}\left(\sqrt{n}\ \overline{\xi}_n\right) = \lim_{n\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{ni}\right)$$

Therefore α^2 is the long-run variance parameter associated to the time series

$$\begin{aligned}\xi_{nk} &= (\mathbf{v}_n'\mathbf{Y}_{nk})(\mathbf{w}_n'\mathbf{Y}_{nk}) - \mathrm{E}\left[(\mathbf{v}_n'\mathbf{Y}_{n1})(\mathbf{w}_n'\mathbf{Y}_{n1})\right]\\ &= \sum_{l_1,l_2=1}^{L} Z^{l_1}_{nk}(\mathbf{v})Z^{l_2}_{nk}(\mathbf{w}) - \mathrm{E}\left[Z^{l_1}_{n1}(\mathbf{v})Z^{l_2}_{n1}(\mathbf{w})\right], \quad k\ge1,\end{aligned}$$

and $\hat{\alpha}_n^2$ is the Bartlett-type estimator calculated from the first n observations with

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

Furthermore we can conclude from the proof, see (A25), that $\beta^2 = \beta^2(r,s,d)$ can be expressed by

$$\beta^2(r,s;d) = \lim_{n\to\infty}\left\{\mathrm{E}[\xi_{n1}(r)\xi_{n1}(s)] + 2\sum_{h=1}^{n}\frac{n-h}{n}\mathrm{E}[\xi_{n1}(r)\xi_{n,1+h}(s)]\right\},$$

where $\xi_{nk}(i) = Y^{(\mathbf{v}_i)}_{nk}Y^{(\mathbf{w}_i)}_{nk} - \mathrm{E}\left[Y^{(\mathbf{v}_i)}_{n1}Y^{(\mathbf{w}_i)}_{n1}\right], k\ge1$, for $i = 1, 2, \dots$.

It suffices to show

$$\sup_{d\in\mathbb{N}} \mathrm{E}|\hat{\beta}_n^2(r,s;d_n) - \beta^2(r,s;d_n)| \overset{n\to\infty}{\to} 0.$$

Define $\Gamma_{nh}(r,s;d) = \mathrm{E}[\xi_{n1}(r)\, \xi_{n,1+|h|}(s)], h \in \mathbb{Z}$. A lengthy explicit calculation using the representation of $\xi_{n1}(r)$ as a finite sum of linear processes shows that

$$\sup_{n\geq 1}\sup_{r,s\geq 1}\sup_{d\in\mathbb{N}} \mathrm{E}[\xi_{n1}(r)\, \xi_{n,1+|h|}(s)] \ll |h|^{-2(1+\delta)},$$

such that

$$\sup_{n\geq 1}\sup_{1\leq r,s}\sup_{d\in\mathbb{N}} |\beta^2(r,s;d)| \leq \sup_{1\leq r,s}\sup_{d\in\mathbb{N}} \sum_{h\in\mathbb{Z}} |\Gamma_h(r,s;d)| < \infty.$$

Wei Biao and Min (2005) introduced the following coupling dependence measure: Let ε_i' be an i.i.d. copy of ε_i. Then $(\dots, \varepsilon_{i-1}, \varepsilon_i')$ is a coupled version of $(\dots, \varepsilon_{i-1}, \varepsilon_i)$ with ε_i replaced by ε_i'. For an arbitrary linear process $Z_n = \sum_{j=0}^{\infty} z_j \varepsilon_{n-j} = Z(\dots, \varepsilon_{n-1}, \varepsilon_n)$ and $Z_n' = Z(\dots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \dots, \varepsilon_n)$ with $\varepsilon_0 \overset{d}{=} \varepsilon_0'$ such that ε_0' is independent from $\{\varepsilon_k\}$, the coupling dependence measure is given by

$$\delta_p(\{Z_i : i \in \mathbb{N}_0\}, n) = \|Z_n - Z_n'\|_{L_p} = |z_n| \|\varepsilon_0 - \varepsilon_0'\|_{L_p}, \quad p \geq 1. \tag{A26}$$

Applying this fact to our situation, we obtain for some constant C for all $n \geq 1$

$$\|Y_{nk}(\mathbf{v})\|_{L_8} \leq \sum_{l=1}^{L} \left\| Z_{nk}^l(\mathbf{v}) \right\|_{L_8} \leq \sum_{l=1}^{L}\sum_{j=0}^{\infty} \left| z_{nj,\mathbf{v}}^{(l)} \right| \underbrace{\left\| \varepsilon_{k-j}^{(l)} \right\|_{L_8}}_{<\infty} \leq C \sum_{j=0}^{\infty} (j \vee 1)^{-(1+\delta)} < \infty$$

and

$$\begin{aligned}\delta_8(\{Y_{ni}(\mathbf{v})\}, k) = \|Y_{nk}(\mathbf{v}) - Y_{nk}(\mathbf{v})'\|_{L_8} &\leq \sum_{l=1}^{L} \left\| Z_{nk}^l(\mathbf{v}) - Z_{nk}^l(\mathbf{v})' \right\|_{L_8} \\ &\overset{(53)}{=} \sum_{l=1}^{L} \left| z_{nk,\mathbf{v}}^{(l)} \right| \left\| \varepsilon_0^{(l)} - \varepsilon_0'^{(l)} \right\|_{L_8} \leq CL(k \vee 1)^{-(1+\delta)},\end{aligned}$$

such that

$$\sup_{n\geq 1} \sup_{\mathbf{v}_n : \|\mathbf{v}_n\|_{\ell_1} \leq C_{v,w}} \sum_{k=0}^{\infty} \delta_8(\{Y_{ni}(\mathbf{v})\}, k) < \infty,$$

that is, uniformly over n and weighting vectors with uniformly bounded ℓ_1-norms. Furthermore,

$$\begin{aligned}\delta_4(\{\xi_{ni}\}, k) &= \|Y_{nk}(\mathbf{v})Y_{nk}(\mathbf{w}) - (Y_{nk}(\mathbf{v})Y_{nk}(\mathbf{w}))'\|_{L_4} \\ &\leq \|Y_{nk}(\mathbf{v})\|_{L_8}\ \|Y_{nk}(\mathbf{w}) - Y_{nk}(\mathbf{w})'\|_{L_8} + \|Y_{nk}(\mathbf{w})\|_{L_8}\ \|Y_{nk}(\mathbf{v}) - Y_{nk}(\mathbf{v})'\|_{L_8} \\ &= O(\delta_8(\{Y_{ni}(\mathbf{v})\}, k) + \delta_8(\{Y_{ni}(\mathbf{w})\}, k)) \\ &= O((k \vee 1)^{-(1+\delta)}),\end{aligned}$$

leading again to $\sup_{n\geq 1}\sup_{\mathbf{v}_n : \|\mathbf{v}_n\|_{\ell_1}\leq C_{v,w}} \sum_{k=0}^{\infty} \delta_4(\{\xi_i\}, k) < \infty$. Finally, similar arguments lead to

$$\delta_2(\{\xi_{ni}(r)\xi_{n,i+h}(s)\}, k) = O(\delta_8(\{Y_{ni}(\mathbf{v})\}; k) + \delta_8(\{Y_{ni}(\mathbf{w})\}; k))$$

and therefore to $\sup_{n\geq 1}\sup_{\mathbf{v}_n : \|\mathbf{v}_n\|_{\ell_1}\leq C_{v,w}} \sum_{k=0}^{\infty} \delta_2(\{\xi_i(r)\ \xi_{i+h}(s)\}, k) < \infty$. Next define $\tilde{\Gamma}_{nh}(r,s) = \frac{1}{n}\sum_{i=1}^{n-h} \xi_{ni}(r)\ \xi_{n,i+h}(s)$. We may employ Wei Biao (2007, th. 1), which applies to the stationary series $\boldsymbol{Y}_{n1}, \dots, \boldsymbol{Y}_{nn}$, and obtain, in view of the fact that the above dependence measures are uniformly bounded in $n\geq 1$ and $r, s\geq 1$,

$$\sup_{d\in\mathbb{N}} \mathrm{E}[n(\tilde{\Gamma}_{nh}(r,s;d) - \mathrm{E}[\tilde{\Gamma}_{nh}(r,s;d)])]^2 \leq C_3(n-h) \tag{A27}$$

for some constant $C_3 < \infty$ not depending on h, d or m, uniformly over $\|\mathbf{v}_n\|_{\ell_1} \leq C_{v,w}$, such that

$$\sup_{1\leq r,s}\sup_{d\in\mathbb{N}}\sup_{|h|\leq m_n} \|\tilde{\Gamma}_{nh}(r,s;d) - \mathrm{E}[\tilde{\Gamma}_{nh}(r,s;d)]\|_{L_2} \leq C_4\ n^{-\frac{1}{2}},$$

for some constant $C_4 < \infty$. With these estimates the proof can be completed as in Steland (2020), see Bours (2019) for details.

A.7 Proof of Theorem 6

By (A15) we have $D_{nk}(\mathbf{v}_n, \mathbf{w}_n) = U_{nk}(\mathbf{v}_n, \mathbf{w}_n) - \mathrm{E}(U_{nk}(\mathbf{v}_n, \mathbf{w}_n))$. In what follows, the weighting vectors $\mathbf{v}_n$ and $\mathbf{w}_n$ are fixed and therefore we omit them in notation. The existence of a $\mathcal{F}_{nk}$-martingale which approximates the D_{nk} under the change-point alternative is crucial and will allow for a straightforward proof of the consistency of the change-point estimator.

Theorem 6 can be shown along the lines of the proof for $L = 1$ given in Steland (2020). For completeness, we provide some details of the proof. The first step is to establish the following martingale approximation under the change-point model, which is of independent interest.

Lemma 6. *Under the change-point model given by (21), (23) and (24) with $\tau = \lfloor n\vartheta \rfloor$, $\vartheta \in (\varepsilon, 1)$ for some $\varepsilon \in (0,1)$, there exist a $\mathcal{F}_{nk}$-martingale array $\tilde{D}_{nk}$, $1\leq k$, $n\geq 1$, such that*

$$\mathrm{E}(D_{nk} - \tilde{D}_{nk})^2 \overset{n,k}{\ll} k^{1-\vartheta}. \tag{A28}$$

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

$$\mathrm{E}\left(D_{nk} - \frac{k}{n}D_{nn} - \left[\tilde{D}_{nk} - \frac{k}{n}\tilde{D}_{nn}\right]\right)^2 \overset{n,k}{\ll} k^{1-\vartheta}. \tag{A29}$$

and for $1\leq k < n$

$$\mathrm{E}\left(\frac{D_{nk}}{k} - \frac{D_{nn} - D_{nk}}{n-k} - \left[\frac{\tilde{D}_{nk}}{k} - \frac{\tilde{D}_{nn} - \tilde{D}_{nk}}{n-k}\right]\right)^2 \overset{n,k}{\ll} n^{-1-\vartheta}. \tag{A30}$$

Proof of Lemma 6. Denote the approximating martingales for D_{nk} corresponding to the coefficients $\mathfrak{b}$ and $\mathfrak{c}$ by $M_k^{(n)}(\mathfrak{b})$ and $M_k^{(n)}(\mathfrak{c})$. For each $n\geq 1$, put

$$\tilde{D}_{nk} = \delta M_k^{(n)}(\mathfrak{b})\mathbf{1}(k\leq\tau) + [\delta M_\tau^{(n)}(\mathfrak{b}) + \delta M_k^{(n)}(\mathfrak{c}) - \delta M_\tau^{(n)}(\mathfrak{c})]\mathbf{1}(k>\tau), \quad k\geq 1. \tag{A31}$$

As in Steland (2020) one verifies that $\mathrm{E}(\tilde{D}_{nk}|\mathcal{F}_{n,k-1}) = 0$ for all k. Let $\boldsymbol{U}_{nk} = (\boldsymbol{U}_{nk}^{(1)}, \boldsymbol{U}_{nk}^{(2)})'$ with $\boldsymbol{U}_{nk}^{(1)} = \sum_{i\le k} \boldsymbol{Y}_{ni}(\mathfrak{b})\boldsymbol{Y}_{ni}(\mathfrak{b}))'$ and $\boldsymbol{U}_{nk}^{(2)} = \sum_{i\le k} \boldsymbol{Y}_{ni}(\mathfrak{c})\boldsymbol{Y}_{ni}(\mathfrak{c}))'$. Clearly,

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

such that the proof of (A29) can be completed as in Steland (2020), as a consequence of Lemma 5 applied to $\mathbf{D}_{nk}^{(1)}$ and $\mathbf{D}_{nk}^{(2)}$. (A30) follows by noting that $\lfloor n\varepsilon \rfloor \le k, n-k \le \lfloor n(1-\varepsilon) \rfloor$, such that, for example,

$$\frac{\mathrm{E}(D_{nk} - \tilde{D}_{nk})^2}{(n-k)^2} \overset{n,k}{\ll} n^{-1-\vartheta}. \qquad \blacksquare$$

We claim that

$$\mathcal{V}_n(k) = \begin{cases} \frac{1-\tau/n}{1-k/n}\Delta_n, & k \le \tau, \\ \frac{\tau/n}{k/n}\Delta_n, & k > \tau. \end{cases}$$

To verify this formula, observe that, by definition of $\hat{\mathcal{V}}_n(k)$

$$\mathrm{E}(\tilde{\mathcal{V}}_n(k)) = \frac{1}{k}\mathrm{E}(U_{nk}) - \frac{1}{n-k}(\mathrm{E}(U_{nn}) - \mathrm{E}(U_{nk})).$$

Plugging in the definition of U_{nk}, we obtain the representation

$$\mathrm{E}(\tilde{\mathcal{V}}_n(k)) = \mathbf{v}_n' \left(\frac{1}{k}\boldsymbol{S}_{nk} - \frac{1}{n-k}(\boldsymbol{S}_{nn} - \boldsymbol{S}_{nk}) \right) \mathbf{w}_n. \tag{A33}$$

We have for $k \le \tau$ by linearity

$$\begin{aligned} \mathrm{E}\left[\frac{1}{k}\boldsymbol{S}_{nk} - \frac{1}{n-k}(\boldsymbol{S}_{nn} - \boldsymbol{S}_{nk})\right] &= \boldsymbol{\Sigma}_n^{(0)} + \frac{1}{n-k}(\tau\boldsymbol{\Sigma}_n^{(0)} + (n-\tau)\boldsymbol{\Sigma}_n^{(1)} - k\boldsymbol{\Sigma}_n^{(0)}) \\ &= \left(1 - \frac{\tau-k}{n-k}\right)\boldsymbol{\Sigma}_n^{(0)} - \frac{n-\tau}{n-k}\boldsymbol{\Sigma}_n^{(1)} \\ &= \frac{n-\tau}{n-k}(\boldsymbol{\Sigma}_n^{(0)} - \boldsymbol{\Sigma}_n^{(1)}) \end{aligned}$$

and for $k > \tau$

$$\begin{aligned} \mathrm{E}\left[\frac{\boldsymbol{S}_{nk}}{k} - \frac{\boldsymbol{S}_{nn} - \boldsymbol{S}_{nk}}{n-k}\right] &= \frac{1}{k}[\tau\boldsymbol{\Sigma}_n^{(0)} + (k-\tau)\boldsymbol{\Sigma}_n^{(1)}] - \frac{1}{n-k}[(n-\tau)\boldsymbol{\Sigma}_n^{(1)} - (k-\tau)\boldsymbol{\Sigma}_n^{(1)}] \\ &= \frac{\tau}{k}\boldsymbol{\Sigma}_n^{(0)} - \left(-\frac{k-\tau}{k} + \frac{n-k}{n-k}\right)\boldsymbol{\Sigma}_n^{(1)} \\ &= \frac{\tau}{k}(\boldsymbol{\Sigma}_n^{(0)} - \boldsymbol{\Sigma}_n^{(1)}). \end{aligned}$$

To summarize

$$\mathrm{E}\left[\frac{\boldsymbol{S}_{nk}}{k} - \frac{\boldsymbol{S}_{nn} - \boldsymbol{S}_{nk}}{n-k}\right] = \begin{cases} \frac{n-\tau}{n-k}(\boldsymbol{\Sigma}_n^{(0)} - \boldsymbol{\Sigma}_n^{(1)}), & k \le \tau, \\ \frac{\tau}{k}(\boldsymbol{\Sigma}_n^{(0)} - \boldsymbol{\Sigma}_n^{(1)}), & k > \tau. \end{cases}$$

In view of (A33) and the definition of Δ_n this implies (22). Introduce the associated rescaled functions

$$\hat{v}_n(t) = \hat{\mathcal{V}}_n(\lfloor nt \rfloor), \quad t \in [\varepsilon, 1-\varepsilon], \tag{A34}$$

$$v_n(t) = \mathcal{V}_n(\lfloor nt \rfloor), \quad t \in [\varepsilon, 1-\varepsilon], \tag{A35}$$

and

$$v(t) = \frac{1-\vartheta}{1-t}\Delta \mathbf{1}(t \leq \vartheta) + \frac{\vartheta}{t}\Delta \mathbf{1}(t > \vartheta), \quad t \in [\varepsilon, 1-\varepsilon]. \tag{A36}$$

We have

$$\tilde{\tau}_n = \underset{\lfloor n\varepsilon \rfloor \leq k \leq \lfloor n(1-\varepsilon) \rfloor}{\arg\max} \hat{v}_n(k/n) = n \underset{t \in \{\lfloor n\varepsilon \rfloor/n, \dots, \lfloor n(1-\varepsilon) \rfloor/n\}}{\arg\max} \hat{v}_n(t) = n\tilde{\vartheta}_n. \tag{A37}$$

The next step is to show that $\hat{v}_n(t)$ converges uniformly to $v(t)$ in the sense that

$$\sup_{t \in [\varepsilon, 1-\varepsilon]} |\hat{v}_n(t) - v(t)| = o_P(1),$$

as $n \to \infty$. First observe that $\hat{\mathcal{V}}_n(k) - \mathcal{V}_n(k) = \left(\frac{\tilde{D}_{nk}}{k} - \frac{\tilde{D}_{nn} - \tilde{D}_{nk}}{n-k}\right) + R_{nk}$, where $\tilde{D}_{nk}$ is the $\mathcal{F}_{nk}$-martingale array from Lemma 6, and, by virtue of (A30), the remainder term $R_{nk} = \left(\frac{D_{nk}}{k} - \frac{D_{nn} - D_{nk}}{n-k} - [\frac{\tilde{D}_{nk}}{k} - \frac{\tilde{D}_{nn} - \tilde{D}_{nk}}{n-k}]\right)$ satisfies

$$E\left(\max_{\lfloor n\varepsilon \rfloor \leq k \leq \lfloor n(1-\varepsilon) \rfloor} R_{nk}^2\right) \overset{n}{\ll} n^{-\vartheta}.$$

Hence, it suffices to show the following maximal inequalities: For all $\delta > 0$

$$P\left(\max_{\lfloor n\varepsilon \rfloor \leq k \leq \lfloor n(1-\varepsilon) \rfloor} |\tilde{D}_{nk}| > \delta n\right) = o(1), \ P\left(\max_{\lfloor n\varepsilon \rfloor \leq k \leq \lfloor n(1-\varepsilon) \rfloor} |\tilde{D}_{nn} - \tilde{D}_{nk}| > \delta n\right) = o(1), \tag{A38}$$

as $n \to \infty$. Using $E(\tilde{D}_{nn}^2) = O(n)$ and Doob's maximal inequality we may conclude that

$$P\left(\max_{\lfloor n\varepsilon \rfloor \leq k \leq \lfloor n(1-\varepsilon) \rfloor} |\tilde{D}_{nk}|^2 > \delta^2 n^2\right) \leq P\left(\max_{\lfloor n\varepsilon \rfloor \leq k \leq n} |\tilde{D}_{nk}|^2 > \delta^2 n^2\right) = \frac{E(\tilde{D}_{nn}^2)}{\delta^2 n^2} = O\left(\frac{1}{n}\right).$$

Furthermore, noting that $\tilde{D}_{nn} - \tilde{D}_{nk}$ is the sum of $n - k = O(n\varepsilon)$ martingale differences, a further application of Doob's inequality shows that

$$P\left(\max_{\lfloor n\varepsilon \rfloor \leq k \leq \lfloor n(1-\varepsilon) \rfloor} |\tilde{D}_{nn} - \tilde{D}_{nk}|^2 > \delta^2 n^2\right) = O\left(\frac{1}{n}\right)$$

as well. Therefore, (A38) follows. Next consider

$$\sup_{t \in [\varepsilon, 1-\varepsilon]} |\hat{v}_n(t) - v(t)| \leq \sup_{t \in [\varepsilon, 1-\varepsilon]} |\hat{v}_n(t) - v_n(t)| + \sup_{t \in [\varepsilon, 1-\varepsilon]} |v_n(t) - v(t)|$$

$$= \max_{\lfloor n\varepsilon \rfloor \le k \le \lfloor n(1-\varepsilon) \rfloor} |\hat{\mathcal{V}}_n(k) - \mathcal{V}_n(k)| + \sup_{t \in [\varepsilon, 1-\varepsilon]} |v_n(t) - v(t)|$$

$$= \sup_{t \in [\varepsilon, 1-\varepsilon]} |v_n(t) - v(t)| + o_{\mathrm{P}}(1),$$

as $n \to \infty$. Observe that $\sup_{t \in [\varepsilon, 1]} |\frac{\lfloor nt \rfloor}{n} - t| = O(1/n)$. Therefore,

$$\sup_{t \in [\varepsilon, 1-\varepsilon]} |v_n(t) - v(t)| \le \sup_{t \in [\varepsilon, \vartheta]} |v_n(t) - v(t)| + \sup_{t \in [\vartheta, 1-\varepsilon]} |v_n(t) - v(t)| = o(1),$$

as $n \to \infty$, and we arrive at $\sup_{t \in [\varepsilon, 1-\varepsilon]} |\hat{v}_n(t) - v(t)| = o_{\mathrm{P}}(1)$. The last step is to apply an argmin theorem: ϑ is an isolated minimum of $-v(\cdot)$ and $-\hat{v}_n(\cdot)$ converges uniformly to $-v(\cdot)$ on the compact set $[\varepsilon, 1-\varepsilon]$, where $0 < \varepsilon < \vartheta$ is arbitrary but fixed. Consequently, we can apply an argmin result, for example, van der Vaart (1998), and now (A37) yields $\tilde{\vartheta}_n \stackrel{\mathrm{P}}{\to} \vartheta > 0$ and therefore the ratio consistency $\tilde{\tau}_n / \tau = \tilde{\vartheta}_n / \vartheta \stackrel{\mathrm{P}}{\to} 1$, as $n \to \infty$.