Submitted to the Annals of Statistics

ONLINE CHANGE-POINT DETECTION FOR MATRIX-VALUED TIME SERIES WITH LATENT TWO-WAY FACTOR STRUCTURE*

By Yong He‡, Xinbing Kong§, Lorenzo Trapani†,¶, and Long Yu‖

Shandong University ‡, Nanjing Audit University §,
University of Leicester ¶, and Shanghai University of Finance and Economics‖

This paper proposes a novel methodology for the online detection of changepoints in the factor structure of large matrix time series. Our approach is based on the well-known fact that, in the presence of a changepoint, the number of spiked eigenvalues in the second moment matrix of the data increases (e.g., in the presence of a change in the loadings, or if a new factor emerges). Based on this, we propose two families of procedures - one based on the fluctuations of partial sums, and one based on extreme value theory - to monitor whether the first non-spiked eigenvalue diverges after a point in time in the monitoring horizon, thereby indicating the presence of a changepoint. Our procedure is based only on rates; at each point in time, we randomise the estimated eigenvalue, thus obtaining a normally distributed sequence which is *i.i.d.* with mean zero under the null of no break, whereas it diverges to positive infinity in the presence of a changepoint. We base our monitoring procedures on such sequence. Extensive simulation studies and empirical analysis justify the theory. An R package implementing the procedure is available on CRAN.*

1. Introduction. In this paper, we focus on developing several online changepoint detection schemes for matrix-valued time series with an (approximate) latent factor structure. To fix ideas, let $\{\mathbf{X}_t, t = 1, 2, \ldots\}$ be a time series of $p_1 \times p_2$ matrices. A matrix factor model with common factors can be written as

$$\underset{p_1\times p_2}{\mathbf{X}_t} = \underset{p_1\times k_1}{\mathbf{R}}\ \underset{k_1\times k_2}{\mathbf{F}_t}\ \underset{k_2\times p_2}{\mathbf{C}'} + \underset{p_1\times p_2}{\mathbf{E}_t}\ ,\ k_1,\ k_2 > 0, \tag{1.1}$$

where the subscripts represent the row and column dimensions of each matrix. In (1.1), $\mathbf{R}$ is a $p_1 \times k_1$ matrix of loadings explaining the variations of $\mathbf{X}_t$ across the rows, $\mathbf{C}$ is a $p_2 \times k_2$ matrix of loadings reflecting the differences across the columns of $\mathbf{X}_t$, $\mathbf{F}_t$ is a matrix of common factors, and $\mathbf{E}_t$ is the idiosyncratic component. In (1.1), the factor numbers k_1 and k_2 are assumed strictly positive, thus demonstrating a collaborative dependence between both the cross-row and the cross-column dimensions; in such a case, the matrix factor structure in (1.1) is also known as a *two-way factor structure*.

A possible approach to analyse matrix-valued time series is to firstly vectorize the data $\mathbf{X}_t$, and then employ the techniques which have been developed for vector factor models. However, when

*Supported by the grant of the National Science Foundation of China (NSFC 12171282, 11801316, 71971118 and 11831008); the Fundamental Research Funds of Shandong University; Young Scholars Program of Shandong University, China; Project funded by China Postdoctoral Science Foundation (2021M701997); the WRJH-QNBJ Project and Qinglan Project of Jiangsu Province.

†Corresponding author.

Keywords and phrases: Matrix factor model; Factor space; Online changepoint detection; Projection estimation; Randomisation.

*https://cran.r-project.org/web/packages/OLCPM/index.html

data genuinely have a two-way factor structure as in (1.1), this approach is bound to lead to sub-optimal inference (see, for example, Wang et al., 2019, and Yu et al., 2022): "[...] analyzing large scale matrix-variate data is still in its infancy, and as a result, scientists frequently analyse matrix-variate observations by separately modelling each dimension or 'flattening' them into vectors. This destroys the intrinsic multi-dimensional structure and misses important patterns in such large scale data with complex structures, and thus leads to sub-optimal results", as quoted from Chen and Fan (2021).

1.1. *Literature review.* In recent years, factor models for matrix-valued time series have been paid significant attention as an alternative to vectorising the data $\mathbf{X}_t$. Wang et al. (2019) - in a similar spirit to Lam and Yao (2012) - propose an estimator of the factor loading matrices (and of numbers of the row and column factors) based on an eigen-analysis of the auto-cross-covariance matrix, under the assumption that the idiosyncratic term $\mathbf{E}_t$ is white noise. Assuming cross-sectional pervasiveness along the row and column dimensions, Chen and Fan (2021) propose an estimation technique for (1.1) based on an eigen-analysis of a weighted average of the mean and the column (row) covariance matrix of the data; Yu et al. (2022) improve the estimation efficiency of the factor loading matrices with iterative projection algorithms (see also He et al., 2023). Extensions of the basic set-up in (1.1) include the constrained version by Chen et al. (2020a), the semiparametric estimators by Chen et al. (2020b), and the estimators developed in Chen et al. (2022) and Han et al. (2022). Applications of (1.1) include dynamic transport networks in the context of international trade flows (Chen and Chen, 2020), and financial data (Chen et al., 2021a).

Conversely, other aspects in the inference on (1.1) are still unexplored. To the best of our knowledge, there is no contribution on determining whether the factor structure in (1.1) is constant over time or not (see however Liu and Chen, 2022, who study in-sample inference in the presence of a threshold structure). This is a crucial question in many applications using matrix-valued data: among the many possible examples, in marketing science (where matrix factor models arise e.g. in the analysis of movie recommendation rating matrix, see Koren et al., 2009), finding evidence of changepoints would help to learn updated consumer preferences; in asset pricing, the presence of a changepoint in series of matrices of several financials recorded for several companies could highlight the presence of e.g. a downside state of the market (see Massacci et al., 2021); in climate studies, detecting changes in an array of environmental variables observed at a group of stations could prove very useful in monitoring pollution (Hopke et al., 2011); and, finally, in the context of physiological time series, matrix factor models can be naturally applied to the analysis of EEG data (where signals are recorded for several patients and several electrodes), and the presence of a changepoint is a useful marker to spot an epileptic manifestation (Lavielle, 2005). In this paper, we are particularly interested in *online detection* of a changepoint, i.e. in the timely detection of a possible break as new data come in. This has important implications on model selection, helping the applied user to amend her/his forecasts as soon as the change is detected.

Changepoint detection in a high-dimensional context has received lots of attention recently, especially in terms of offline/in-sample detection - see e.g. the recent contributions by Wang and Samworth (2018), Chen et al. (2020c) and Chen et al. (2021b). As far as detection of changes in the second-order structure of high dimensional vectors is concerned, in-sample changepoint detection is well-developed in the context of vector factor models: examples include Breitung and Eickmeier (2011), Chen et al. (2014), Corradi and Swanson (2014), Han and Inoue (2015), Yamamoto and Tanaka (2015), Cheng et al. (2016), Baltagi et al. (2017), Barigozzi et al. (2018), and Baltagi et al. (2021). Bai et al. (2022) study in-sample detection based on the CUSUM process of the estimated factors, arguing that a changepoint in the factor structure is equivalent to a change in the second

moment of the common factors; this approach ensures that the CUSUM process is low-dimensional, and therefore it can be studied with the standard asymptotic theory developed e.g. in Csörgő and Horváth (1997), but it does not exploit the "blessing of dimensionality" which characterizes factor models (Li et al., 2018), and can boost the changepoint detection ability. Moreover, these approaches have in common the maintained assumption that the covariance matrix of the common factors does not change: when this is not the case (i.e. in the presence of heteroskedastic factors), then a changepoint may be detected even if loadings do not change, which may be viewed as a limitation (see Massacci, 2021, and Wang, 2023).

Barigozzi and Trapani (2020) develop a procedure specifically aimed at sequential changepoint detection, exploiting the large dimension of the dataset, for vector-valued time series. Hence, a possible approach to detect changes in a matrix factor model could be based on vectorising the data $\mathbf{X}_t$, and then applying the techniques which have been developed for the changepoint analysis of vector factor models. However, such an approach would be subject to the criticisms mentioned above, since it ignores the spatial relationship of an array - for example, in the context of movie recommendation rating matrices, customers with certain common characteristics tend to favour some featured movies, and this would be missed by a vector factor model.

1.2. *Contributions and structure of the paper.* In this paper, we develop a procedure specifically designed for the online detection of changepoints in a matrix factor model. In particular, we study the detection of changes in the row and column factor spaces, spanned by the columns of $\mathbf{R}$ and $\mathbf{C}$ in (1.1). Our methodology is entirely novel, both in the context of matrix factor models (where we are not aware of any contributions on this issue) and in the more general context of sequential detection (for which we develop a completely original approach).

Hereafter, we provide a heuristic description of how our methodology works, and list our contributions to the current literature on matrix factor models and on sequential detection. Further details are in the remainder of the paper. We focus on two leading examples of possible changes: *(i)* the case where the loadings associated with the common (row) factors change, and *(ii)* the case where a new factor (or several new factors) emerges, thereby enlarging the factor space. Further alternative hypotheses, including e.g. the case of vanishing factors (thus corresponding to a contraction of the factor spaces), can also be studied, in principle, using our approach (see Section B.1.1). In all cases, we base our methodology on two facts which are well-known in the context of vector factor models. First, in a model with a common factor structure, the second moment matrix of the data has a spiked spectrum: in the presence of, say, k common factors, the first k eigenvalues diverge with the sample size, whereas the others are bounded. Second, in the presence of a break, the number of common factors increases across the changepoint: a factor model with a changepoint can be written equivalently as a factor model with no changepoint and an enlarged factor space (the second moment matrix of the data, consequently, having more spiked eigenvalues). The former result is well-known throughout the literature, and in the context of matrix factor models (see e.g. He et al., 2023): in this paper we further extend and refine the results in He et al. (2023) by considering the L_r-norms of the eigenvalues of the second moment matrix of the data. The latter result is also well-known in the context of vector factor models (see e.g. Corradi and Swanson, 2014); in this paper, we extend it to the matrix factor model context, also developing *ad-hoc* methodologies for this type of datasets.

On account of the considerations above, we propose an online monitoring scheme based on sequentially monitoring the eigenvalues of the second moment matrix of the data after a training period in which no break occurs. In particular if, during the training period, k common factors have been found, we monitor the $(k+1)$-th eigenvalue: if no break occurs, this will be bounded

over the monitoring horizon, otherwise it will have a spike at some point. Monitoring the $(k+1)$-th eigenvalue has therefore the advantage that the (large) cross-sectional dimension is fully exploited through the spiked structure (or lack thereof) of the spectrum; in principle, it would be possible to use different approaches which would turn the problem into a low-dimensional one (as e.g. in Bai et al., 2022), but this involves missing the opportunity of exploiting the high-dimensionality of the dataset. We are not aware of any results on the second-order asymptotics of the estimated eigenvalues, especially when these are non-spiked: indeed, as the results in Section 2 show, the estimated $(k+1)$-th eigenvalue may not even be consistent, and may pass to positive infinity even though the corresponding population eigenvalue is bounded. Hence, we do not attempt to derive the asymptotic distribution of the estimated eigenvalue used for the monitoring scheme. Instead, we propose a monitoring scheme based only on *rates*.

In particular, we propose a transformation of the estimated $(k+1)$-th eigenvalue, which is designed to drift to zero under the null of no change (when the $(k+1)$-th population eigenvalue is not spiked), and to pass to infinity after a change (when the $(k+1)$-th population eigenvalue becomes spiked). Such a transformation is based on the L_r-norm of the $(k+1)$-th eigenvalue, and it is designed so as to boost both rates under the null and under the alternative; this transformation is data-driven, and, as a by-product, we discuss how to find r in practice. In principle, the sequence of the L_r-norm of the $(k+1)$-th eigenvalues estimated over time could be used directly: at each point in time during the monitoring horizon, the L_r-norm could be contrasted against a pre-specified threshold, thus indicating whether a new factor has emerged (and, therefore, a changepoint has occurred) or not (thus indicating that no break has occurred in the factor structure). Such an approach would be related (and similar in spirit) to the estimation of the number of common factors via eigenvalue thresholding which is often employed in the (vector) factor model literature (see e.g. Bai and Ng, 2002 as a prime example). However, using a pre-specified threshold is bound to have an element of arbitrariness, and therefore such an approach would be subject to the same criticism as in Hallin and Liška (2007).

Hence, in order to avoid having to set a somewhat arbitrary threshold, we construct a full-fledged sequential testing procedure, where the transformed $(k+1)$-th eigenvalue is contrasted against a critical value, which guarantees controllable Type I and Type II errors. In order to design a test, and in the absence of a limiting distribution for the transformed $(k+1)$-th eigenvalue, we propose a *randomised* testing procedure. Using randomisation in order to have a workable limiting distribution has recently acquired popularity in econometrics and statistics, and it has been shown to be helpful in a wide variety of non-standard inferential problems, chiefly when a limiting distribution is not available or depends on nuisance parameters. Examples of applications include, *inter alia*, tests for forecasting ability (Corradi and Swanson, 2006) and nonstationarity (Bandi and Corradi, 2014), testing for the existence of the moments of a random variable (Trapani, 2016), testing when a parameter is on the boundary under the null (Horváth and Trapani, 2019), estimating the number of common factors in a vector or matrix factor models (Trapani, 2018, He et al., 2023), and estimation in the presence of cointegration and heavy tails (Barigozzi et al., 2022). As mentioned above, Barigozzi and Trapani (2020) use randomisation in the context of changepoint sequential detection in a vector factor model. To the best of our knowledge, this contribution is the first and only attempt to carry out sequential changepoint detection in a vector factor model; however, the approach proposed therein depends on many tuning parameters (see also Section 3 for a more specific description).

Hence, we propose a completely novel, much simpler, virtually tuning-free, approach, which is the first contribution to study online changepoint detection in a factor model for matrix-valued

data. Details are in Section 3; here, we offer a heuristic preview of how our methodology works. At each point in time during the monitoring horizon, we compute the $(k+1)$-th eigenvalue of the second moment matrix of the data, and subsequently perturb its L_r-norm by adding a randomly generated *i.i.d.* sequence with common standard normal distribution. Hence, under the null of no change, such a sequence remains (asymptotically, and conditionally on the sample) an *i.i.d.* sequence with common standard normal distribution; conversely, under the alternative of a changepoint, such a sequence has a diverging mean after the changepoint. This randomisation scheme has several advantages. Firstly, it is immediate to use this sequence to construct a family of weighted CUSUM statistics, essentially based on using the fluctuations of (weighted or unweighted) partial sums, building on the same approach as in Chu et al. (1996), Horváth et al. (2004), and Horváth et al. (2007) - the fact that our sequence is, conditionally on the sample, *i.i.d.* with common standard normal distribution entails that the limiting distributions which we employ are bound to offer very good approximations to the finite sample behaviour of our statistics. Indeed, as a further contribution made possible by using randomisation, we are also able to propose a completely novel monitoring approach, based on taking the maximum value of the perturbed sequence: conditionally on the sample, this is tantamount to using the maximum of an *i.i.d.* sequence with common standard normal distribution, which is a well-known and well-studied problem in statistics. Without randomisation, even if the limiting distribution of the $(k+1)$-th largest eigenvalues were known, these would form a sequence whose dependence structure would be hard to study and would affect the use of Extreme Value Theory. Secondly, using a procedure based only on rates offers another important advantage in the construction of the monitoring procedure. The "classical" approach in the literature (see e.g. Chu et al., 1996) is based on constructing a *detector* where, in essence, an estimator during the monitoring horizon is compared to an estimator based on the training sample: the difference between the two forms a CUSUM-type statistic, whose fluctuations, under the null, never exceed a threshold, or *boundary function*; detection takes place when the boundary function is exceeded. A well-known issue in this context is how to specify the boundary function: its choice is not unique, being, as Chu et al. (1996) put it, "often dictated by mathematical convenience rather than optimality" (p. 1052). Conversely, in our case we know that, by construction, the estimated $(k+1)$-th largest eigenvalues during the training sample drift to zero. Thus, we construct our detector based only on the sequence of the $(k+1)$-th largest eigenvalues estimated during the monitoring horizon; in turn, this makes the construction of the boundary function entirely natural, and based on well-known results on the fluctuations of partial sums. Thirdly, the sequence of the randomised $(k+1)$-th largest eigenvalue is (asymptotically and conditional on the sample) independent with unit variance by construction: therefore, we do not need to estimate any asymptotic variance in order to standardise our statistics, which avoids the (vexing) estimation of a nuisance parameter (see e.g. Müller, 2013, on this issue). As a final remark, using the eigenvalues of the second moment matrix of the data does not require the common factors to be homoskedastic: in the presence of a change in the covariance matrix of the common factors, but without a change in the loadings, our procedures would not find a changepoint. Such robustness is desirable e.g. in macroeconomics and finance, where it is important to discern between changes in the response to the common factors (corresponding to a change in the loadings) and changes in the variance of the factors, and we refer to Massacci (2021), and Wang (2023) for discussion and examples. All these advantages are a consequence of using the eigenvalues of the second moment matrix and of our randomisation scheme and do not depend on the specific model studied herein: thus, our paper also offers a methodological contribution to the changepoint literature in general.

Finally, we would like to point out that, in order to develop our procedure, we make full use

of the matrix structure of the data. Indeed, in our context, it is important to use methods that are specifically designed for matrix-valued time series: as the quote from Chen and Fan (2021) above shows, and as also confirmed by our simulations, using methods based on vectorising the data and subsequently apply the theory designed for vector-valued time series is possible, but it yields sub-optimal results. As a by-product, we develop several technical results, extending the current results on the spectra of second moment matrices of matrix-valued data with a factor structure. In particular, we complement the results in He et al. (2023) - who develop a.s. rates for the estimated eigenvalues - by studying the a.s. behaviour of the L_r-norms thereof. These results are of independent interest.

The remainder of the paper is organized as follows. Sections 2 and 3 contain the main theoretical and methodological results. Implementation guidelines are discussed, and simulations are conducted, in Section 4. An illustrative empirical study is carried out in Section 5; Section 6 concludes, and it also contains a list of all extensions that are reported in the Supplement. Extensions, further results, assumptions and derivations, as well as further simulations an empirical illustrations, are relegated to the Supplement.

NOTATION. Recall that the (row and column) dimensions of $\mathbf{X}_t$ are p_1 and p_2; we will denote the training (pre-monitoring) sample size as m. Given an $n_1 \times n_2$ matrix $\mathbf{A}$, we denote its transpose as $\mathbf{A}'$; its Euclidean and Frobenius norm as $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ respectively; and its element in position (i,j) as A_{ij}, with $1 \leq i \leq n_1$ and $1 \leq j \leq n_2$; and we denote its j-th column as $\mathbf{A}_{\cdot j}$ and its i-th row as $\mathbf{A}_{i\cdot}$. When $n_1 = n_2$, we denote the eigenvalues of $\mathbf{A}$ as $\lambda_i(\mathbf{A})$, sorted in decreasing order. Identity matrices of order n are denoted as I_n. Throughout the paper, we often use the following sequence

$$l_{p_1,p_2,m} = \left(\frac{1}{p_2} + \frac{1}{m} + \frac{p_1}{\sqrt{mp_2}}\right)\left(\ln^2 p_1 \ln p_2 \ln m\right)^{1+\epsilon}. \tag{1.2}$$

Positive, finite constants are denoted as c_0, c_1, ..., and their values may change from line to line. Throughout the paper, we use the short-hand notation "a.s." for "almost sure(ly)". Given two sequences $a_{p_1,p_2,T}$ and $b_{p_1,p_2,T}$, we say that $a_{p_1,p_2,T} = o_{a.s.}(b_{p_1,p_2,T})$ if, as $\min\{p_1,p_2,T\} \to \infty$, it holds that $a_{p_1,p_2,T} b^{-1}_{p_1,p_2,T} \to 0$ a.s.; we say that $a_{p_1,p_2,T} = O_{a.s.}(b_{p_1,p_2,T})$ to denote that as $\min\{p_1,p_2,T\} \to \infty$, it holds that $a_{p_1,p_2,T} b^{-1}_{p_1,p_2,T} \to c_0 < \infty$ a.s.; similarly, we use the notation $a_{p_1,p_2,T} = \Omega_{a.s.}(b_{p_1,p_2,T})$ to indicate that, as $\min\{p_1,p_2,T\} \to \infty$, it holds that $a_{p_1,p_2,T} b^{-1}_{p_1,p_2,T} \to c_0 \in (0,\infty)$.

2. Spectra preliminaries. In this section, we briefly report some preliminary results on the spectra of the row (column) space of the second moment matrix of the data. The rationale underpinning our monitoring schemes is based on these results, taken from He et al. (2023). For the sake of a concise discussion, we relegate the required assumptions (and their discussion) to Section A of the Supplement. Here, we note that we make more explicit the moment restrictions in He et al. (2023), allowing for $E\|\mathbf{F}_t\|^{4r} < \infty$ and $E|e_{ij,t}|^{4r} < \infty$, for some $r \geq 2$ and explicitly using r in the construction of our test statistics - in comparison, the paper by He et al. (2023) requires $r = 2$, but the fact that r may be larger than 2 is not exploited. Further, their results require "strong factors", i.e. they require that $p_1^{-1}\mathbf{R}'\mathbf{R}$ and $p_2^{-1}\mathbf{C}'\mathbf{C}$ converge, as $p_1 \to \infty$ and $p_2 \to \infty$ respectively, to positive definite matrices. In this respect, we note that, similarly to He et al. (2023), we have used the conventions that $\|\mathbf{R}\|_F = c_0 p_1^{1/2}$ and $\|\mathbf{C}\|_F = c_1 p_2^{1/2}$ - i.e., we have embedded the impact of the dimensions p_1 and p_2 into the matrices $\mathbf{R}$ and $\mathbf{C}$ respectively, instead of into $\mathbf{F}_t$. In Section B.1.3 in the Supplement, we discuss the case where factors are weak, i.e. the case $\|\mathbf{R}\|_F = c_0 p_1^{\alpha/2}$ with $\alpha < 1$.

Recall the two-way factor model (1.1)

$$\underset{p_1\times p_2}{\mathbf{X}_t} = \underset{p_1\times k_1}{\mathbf{R}}\,\underset{k_1\times k_2}{\mathbf{F}_t}\,\underset{k_2\times p_2}{\mathbf{C}'} + \underset{p_1\times p_2}{\mathbf{E}_t}\ ,\ k_1,\ k_2 > 0,$$

where, as mentioned above, we focus on the case $k_1 > 0$ and $k_2 > 0$ - the case where either $k_i = 0$ (or both), also investigated in He et al. (2023), can be studied with essentially the same approach, and we briefly discuss this in Section B.1.2 in the Supplement. In the presentation hereafter, we focus on detecting a changepoint in the *row* factor structure of $\mathbf{X}_t$ - thus, the focus is on $\mathbf{R}$, although all the theory developed hereafter can be readily adapted to check for changes in $\mathbf{C}$.

In order to fully make use of the two-way interactive factor structure, we propose studying the spectrum of a projected column (row) covariance matrix, as suggested by Yu et al. (2022). Heuristically, if $\mathbf{C}$ is known and satisfies the orthogonality condition $\mathbf{C}'\mathbf{C}/p_2 = I_{k_2}$, the data matrix can be projected into a lower dimensional space by setting $\mathbf{Y}_t = \mathbf{X}_t\mathbf{C}/p_2$. In view of this, we define

$$\widehat{\mathbf{M}}_1 = \frac{1}{m}\sum_{t=1}^{m}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t', \tag{2.1}$$

where $\widetilde{\mathbf{Y}}_t = p_2^{-1}\mathbf{X}_t\widetilde{\mathbf{C}}$ and $\widetilde{\mathbf{C}}$ is an estimator of $\mathbf{C}$. As suggested by Yu et al. (2022), this can be set as $\widetilde{\mathbf{C}} = \sqrt{p_2}\mathbf{Q}$, where the columns of $\mathbf{Q}$ are the leading, orthonormalised k_2 eigenvectors of $\mathbf{M}_r$, where $\mathbf{M}_r$ is the column "flattened" sample covariance matrices, i.e.,

$$\mathbf{M}_r := \frac{1}{Tp_1}\sum_{t=1}^{T}\mathbf{X}_t'\mathbf{X}_t = \frac{1}{Tp_1}\sum_{t=1}^{T}\sum_{j=1}^{p_1}\mathbf{X}_{j\cdot,t}\mathbf{X}_{j\cdot,t}'. \tag{2.2}$$

Henceforth, we assume that the number of common row and column factors, k_1 and k_2, are known. In Section B.2.1, we study the case where k_1 and k_2 need to be estimated, showing that all our results hold upon using a consistent estimator of k_1 and k_2 and also briefly describing an estimator of k_1 and k_2.

Let $\widehat{\lambda}_j$ denote the j-th largest eigenvalue of $\widehat{\mathbf{M}}_1$. The following result, shown in Theorem 2 of He et al. (2023), quantifies the eigen-gap of $\widehat{\mathbf{M}}_1$.

LEMMA 1. *We assume that Assumptions A.1-A.4 in Section A of the Supplement are satisfied, and that $k_1 > 0$ and $k_2 > 0$. Then it holds that*

$$\widehat{\lambda}_j = \Omega_{a.s.}\left(p_1\right), \tag{2.3}$$

for all $j \leq k_1$. Also, for all $j > k_1$ it holds that

$$\widehat{\lambda}_j = o_{a.s.}\left(l_{p_1,p_2,m}\right), \tag{2.4}$$

for all $\epsilon > 0$, where $l_{p_1,p_2,m}$ is defined in (1.2).

Lemma 1 is shown in He et al. (2023), and it states that there exists an eigen-gap between the largest k_1 eigenvalues and the remaining ones: whilst the former diverge at a rate p_1, the remaining ones are of lower order. The spectrum of $\widehat{\mathbf{M}}_1$ can be compared with the spectrum of $\mathbf{M}_r$, which is

studied in Theorem 1 in He et al. (2023): in that case, the eigen-gap is narrower, thus suggesting that the use of $\widehat{\mathbf{M}}_1$ should yield better results than $\mathbf{M}_r$. As mentioned above, in the statement of the lemma we only focus on the case $k_1 > 0$ and $k_2 > 0$. In He et al. (2023), both the eigen-gaps of $\widehat{\mathbf{M}}_1$ and $\mathbf{M}_r$ are studied also for the case $k_i = 0$ for $i = 1$ and/or 2; the results in He et al. (2023) form the basis for our discussion of this case in Section B.1.2 in the Supplement.

3. Online changepoint detection. Recall that we assume that we have observed data over a training sample $1 \leq t \leq m$, and that no breaks were found, viz.

ASSUMPTION 1. *The column space of* $\mathbf{R}$ *does not change during* $1 \leq t \leq m$.

Assumption 1 can be verified, and k_1 can be computed, with the methodologies described in Yu et al. (2022) and He et al. (2023).

We monitor for changepoints in the row factor structure of $\mathbf{X}_t$ across an interval $m+1 \leq t \leq m+T_m$. Henceforth, T_m represents the size of the monitoring horizon; further, tidying up the notation, we define $T = m + T_m$. The monitoring schemes are based on

$$\widehat{\lambda}_{k_1+1,\tau} = \lambda_{k_1+1}\left(\frac{1}{m}\sum_{t=\tau+1}^{m+\tau}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t'\right).$$

We now propose a method for the sequential detection of two types of breaks; we discuss further alternatives in Section B.1.1. All alternative hypotheses under consideration are nested in

$$\mathbf{X}_t = \begin{cases} \mathbf{R}\mathbf{F}_{1,t}\mathbf{C}' + \mathbf{E}_t & \text{for } 1 \leq t \leq m + t^* \\ \widetilde{\mathbf{R}}\mathbf{F}_{2,t}\mathbf{C}' + \mathbf{E}_t & \text{for } t > m + t^* \end{cases}, \tag{3.1}$$

where $t^* < T_m$.

A first source of change could be a scenario in which the space spanned by the common factors is constant across regimes, but the row factor space spanned by the columns of $\mathbf{R}$ changes after a point in time t^* with

$$H_{A,1} : \mathbf{R} = [\mathbf{R}_0|\mathbf{R}_1] \quad \text{and } \widetilde{\mathbf{R}} = [\mathbf{R}_0|\mathbf{R}_2], \tag{3.2}$$

where $\mathbf{R}_0$ is a $p_1 \times (k_1 - c_1)$ matrix of loadings which does not undergo a change, and $\mathbf{R}_1$ and $\mathbf{R}_2$ are $p_1 \times c_1$ matrices of loadings which differ before and after the changepoint t^*, for some $c_1 \geq 1$. We refer to Assumption A.5 in the Supplement (and to the discussion thereafter) for details on this case; here, we point out that our methodology is robust to the scenario where $\mathbf{R}$ and $\widetilde{\mathbf{R}}$ span the same space, which is tantamount to a matrix factor model where the common factors have a change in volatility and the loadings do not change. Under (3.1), Lemma C.3*(i)* in the Supplement yields

$$\widehat{\lambda}_{k_1+1,\tau}\begin{cases} \leq c_0 l_{p_1,p_2,m} & \text{for } \tau \leq t^* \\ \geq c_1 \min\left\{\frac{\tau - t^*}{m}, \frac{m+t^*-\tau}{m}\right\} p_1 & \text{for } t^* < \tau < m + t^* \\ \leq c_0 l_{p_1,p_2,m} & \text{for } \tau \geq m + t^* \end{cases}. \tag{3.3}$$

Heuristically, this alternative hypothesis is similar to the so-called "epidemic" alternative studied in the context of in-sample changepoint detection.

As a second possible alternative, we consider the scenario whereby a set of common factors appear after the point in time t^*, i.e., the column space of $\mathbf{R}$ becomes larger - in this case

$$H_{A,2} : \widetilde{\mathbf{R}} = [\mathbf{R}|\mathbf{R}_3]\,, \tag{3.4}$$

where $\mathbf{R}_3$ is a $p_1 \times c_3$ matrix, and $\mathbf{F}'_{2,t} = \left[\mathbf{F}'_{1,t}|\widetilde{\mathbf{F}}'_{2,t}\right]$ with $\mathbf{F}'_{2,t}$ is a $k_2 \times c_3$ matrix of new common factors with $c_3 \geq 1$. Lemma C.3*(ii)* in the Supplement stipulates that, in such a case

$$\widehat{\lambda}_{k_1+1,\tau} \begin{cases} \leq c_0 l_{p_1,p_2,m} & \text{for } \tau \leq t^* \\ \geq c_1 \frac{\tau - t^*}{m} p_1 & \text{for } t^* < \tau < m + t^* \\ \geq c_0 p_1 & \text{for } \tau \geq m + t^* \end{cases} . \tag{3.5}$$

This alternative hypothesis, in contrast with (3.3), implies a lasting effect in the eigenvalue $\widehat{\lambda}_{k_1+1,\tau}$.

As a final remark, we point out that in all the cases above we rely on the maintained assumption that $\mathbf{C}$ is constant. We investigate the impact of changes in $\mathbf{C}$ on the size and power of our procedures in Section B.1.4 in the Supplement.

Based on (3.3) and (3.5), we propose a procedure for the null hypothesis of no changepoint over the monitoring horizon in (1.1). We begin by offering a heuristic preview of the rationale underpinning our approach. Under H_0, Lemma 1 implies that, for sufficiently large m, p_1 and p_2

$$\widehat{\lambda}_{k_1+1,\tau} \leq c_0 l_{p_1,p_2,m}, \tag{3.6}$$

for all $1 \leq \tau \leq T_m$. Under either alternative hypothesis, at some point in time $\tau > t^*$, it will hold that, for some positive c_0

$$\widehat{\lambda}_{k_1+1,\tau} \geq c_0 p_1. \tag{3.7}$$

Define now $0 \leq \delta < 1$, such that

$$p_1^{-\delta} \widehat{\lambda}_{k_1+1,\tau} \begin{cases} o_{a.s.}(1) & \text{under no break} \\ \geq c_0 p_1^{1-\delta} \to \infty & \text{if there is a break} \end{cases} . \tag{3.8}$$

The effect of this transformation is that $p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}$ drifts to zero or infinity according as the null or the alternative is true: any δ which satisfies this is, in principle, admissible. Based on (3.6) and (3.7), $p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau} = o_{a.s.}(1)$ as long as

$$\begin{cases} \delta = \varepsilon & \text{if } \beta \leq 1/2 \\ \delta = 1 - 1/(2\beta) + \varepsilon & \text{if } \beta > 1/2 \end{cases} , \tag{3.9}$$

where $\beta = \ln p_1 / \ln(p_2 m)$, and $\varepsilon > 0$ is an arbitrarily small, user-defined number. We discuss the choice of ε, and analyse the robustness of our methodology to such choice, in Section 4. Here, we point out that ε is the only tuning parameter specifically needed by our procedure. The requirement $\varepsilon > 0$ is needed in order to have $p_1^{-\varepsilon}\left(\ln^2 p_1 \ln p_2 \ln m\right)^{1+\epsilon} \to 0$, i.e. to smooth away the slowly varying sequence in (1.2). In the light of this, we need the following assumption.

ASSUMPTION 2. *It holds that*

$$p_1 = O\left((\min\{m, p_2\})^c\right),$$

for some $c < \infty$.

Assumption 2 states that the sample size p_1 can be much larger, but not "too large", compared to m and p_2; in essence, the assumption rules out that p_1 can grow exponentially with m and p_2. Indeed, looking at the definition of δ in (3.9), if p_1 were as large as $\exp\{p_2\}$ or $\exp\{m\}$, δ would have to be equal to 1, which would entail smoothing away even the spiked eigenvalues.

The dichotomous rate behaviour in (3.8) is the building block of our monitoring schemes, and it is further enhanced by using the following transformation

$$\psi_\tau = \left| \frac{p_1^{-\delta} \widehat{\lambda}_{k_1+1,\tau}}{p_1^{-1} \sum_{j=1}^{p_1} \widehat{\lambda}_{j,\tau}} \right|^r . \tag{3.10}$$

In (3.10), r is defined in Assumptions A.1 and A.3 in the Supplement such that $E\|\mathbf{F}_t\|^{4r} < \infty$ and $E|e_{ij,t}|^{4r} < \infty$ with $r \geq 2$. The transformation in (3.10) is "prescriptive": given r, the definition of ψ_τ follows automatically. By (3.8), it follows that $\psi_\tau = o_{a.s.}(1)$ for all $1 \leq \tau \leq T_m$ under the null of no changepoint; conversely, in the presence of a changepoint in t^* it holds that $\psi_\tau \overset{a.s.}{\to} \infty$ for some $\tau > t^*$. In (3.10), $\widehat{\lambda}_{k_1+1,\tau}$ is divided by the trace of $m^{-1}\sum_{t=\tau+1}^{m+\tau} \widetilde{\mathbf{Y}}_t \widetilde{\mathbf{Y}}_t'$ so as to make it scale-invariant; in principle, other rescaling schemes are possible, but the one used here is the one typically employed in this literature (see, *inter alia*, Trapani, 2018, Barigozzi and Trapani, 2020, Barigozzi and Trapani, 2022 and He et al., 2023).

Finally, after computing ψ_τ at each $1 \leq \tau \leq T_m$, we define the sequence $\{y_\tau, 1 \leq \tau \leq T_m\}$ as

$$y_\tau = z_\tau + \psi_\tau, \tag{3.11}$$

where $z_\tau \sim i.i.d.N(0,1)$ for $1 \leq \tau \leq T_m$. We can now add to the comments made in the introduction. As mentioned, the sequence ψ_τ could be employed directly for changepoint detection, by contrasting it against a pre-specified, user-defined threshold at each $1 \leq \tau \leq T_m$. In order to avoid having to deal with the arbitrariness that such a choice would inevitably have, we instead construct the sequence $\{y_\tau, 1 \leq \tau \leq T_m\}$, and we use it to compute the "traditional" test statistics employed in this context, such as the CUSUM-based ones discussed in Section 3.1.1 below. Using ψ_τ (or, for that matter, $\widehat{\lambda}_{k_1+1,\tau}$) directly for this purpose is not possible, since we do not have a limiting distribution for $\widehat{\lambda}_{k_1+1,\tau}$ (and, therefore, *a fortiori*, for ψ_τ). The randomisation scheme defined in (3.11) has the added bonus that the sequence $\{y_\tau, 1 \leq \tau \leq T_m\}$ is *i.i.d.* and Gaussian conditional on the sample, and therefore the asymptotics is easily tractable.

We conclude our discussion by adding a more in-depth comparison with the approach used in Barigozzi and Trapani (2020). In that paper, similarly to ours, monitoring is applied to a transformation of $\widehat{\lambda}_{k_1+1,\tau}$ - although, in Barigozzi and Trapani (2020), such a transformation is not data-driven and its choice itself is a matter of tuning. At each $1 \leq \tau \leq T_m$, such a transformation is then randomised *twice*. The first time,[1] it is: *(a)* randomised by multiplying it by a B-dimensional sequence of *i.i.d.* standard normal variables, where the sample size B is a user-defined tuning parameter; *(b)* transformed into a B-dimensional Bernoulli sequence whose elements are equal to zero or one according as the corresponding elements of the previous sequence exceeds a user-defined threshold or not; and, *(c)* averaged across B. The output of this is a single random variable for each τ under the alternative that there is a change, and a random variable passing to infinity under the null of no change. Then, at the second randomisation round, the reciprocal of such a random

[1]The randomisation scheme used in Barigozzi and Trapani (2020) is essentially the same as the one described in Section B.2.1.

variable are randomised, repeating, for each τ, the procedure above. Hence, an *i.i.d.* sequence obtains, which, as the sample size (and also the size of the artificial, twice-added, random sample B) passes to infinity, is $N(0,1)$ distributed under the null of no change. CUSUM statistics are then constructed from this sequence. Whilst this approach could, in principle, be used here as well, it is laden with tuning parameters in addition to the value of ε in (3.9): the transformation to be applied to $\widehat{\lambda}_{k_1+1,\tau}$ (twice); the artificial sample sizes B in the two randomisations; the threshold to be used in both randomisations to construct the Bernoulli sequences; furthermore, the resulting sequence of randomised, transformed eigenvalues is *i.i.d.* Gaussian only asymptotically (conditionally on the sample), whereas in our context it is exactly *i.i.d* Gaussian (conditional on the sample). All these tuning parameters are avoided when using the randomisation scheme proposed in (3.11).

3.1. *Monitoring schemes.* Sequential monitoring can be carried out in two ways, both based on (3.11): a more traditional one using the fluctuations of partial sums (Section 3.1.1), and an entirely novel one based on the worst case scenario across the monitoring horizon (Section 3.1.2).

In both cases, we allow the monitoring T_m to go on for a long time.

ASSUMPTION 3. *It holds that: (i)* $T_m = \Omega(m)$*; and (ii)*

$$\lim_{\min\{p_1,p_2,m\}\to\infty} T_m \left| p_1^{-\delta} l_{p_1,p_2,m} \right|^r = 0. \tag{3.12}$$

Assumption 3*(i)* is similar e.g .to equation (1.12) in Horváth et al. (2007). In principle, the monitoring horizon could even be much bigger than m, i.e. allowing for $T_m = \Omega(m^\varsigma)$ with $\varsigma \geq 1$ user-chosen, and we discuss this case in Section B.2.2 in the Supplement. Part *(ii)* of the assumption is required to control for a non-centrality parameter which arises in the asymptotics of the test statistic under H_0.

We now present our proposed monitoring schemes. For each of them, we study the limiting distribution under the null, investigate power against alternatives, and, finally, we discuss the delay in changepoint detection. This discussion forms the basis of the implementation guidelines reported in Section 4.1.

As far as the notation is concerned, henceforth we let P^* denote the probability conditional on $\{\mathbf{X}_t, 1 \leq t \leq T\}$; we use "$\xrightarrow{P^*}$", and "$\xrightarrow{D^*}$" to denote convergence in probability and in distribution according to P^*, respectively.

3.1.1. *Monitoring schemes based on partial sum processes.* We begin with a family of "traditional" monitoring schemes, based on partial sum processes. Recall the definition of y_τ in (3.11). We base our test on the partial sums process

$$S_\tau = \sum_{j=0}^{\tau} y_j. \tag{3.13}$$

We will use:

(i) the *weighted functionals*

$$T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta}, \tag{3.14}$$

for $0 \leq \eta < 1/2$;

(ii) the *standardised partial sums*

$$\max_{1\leq\tau\leq T_m}\frac{|S_\tau|}{\tau^{1/2}}; \tag{3.15}$$

(iii) the *Rényi statistics* (see Horváth et al., 2020)

$$r_{T_m}^{\eta-1/2}\max_{r_{T_m}\leq\tau\leq T_m}\frac{|S_\tau|}{\tau^{\eta}}, \tag{3.16}$$

for $\eta > 1/2$.

In (3.16), r_{T_m} is a sequence such that, as $T_m \to \infty$,

$$r_{T_m}\to\infty \quad \text{and} \quad \frac{r_{T_m}}{T_m}\to 0. \tag{3.17}$$

In (3.14)-(3.16), we use S_τ. Typically, when using the fluctuations of partial sums in order to detect changepoints, the parameter to be monitored is estimated at each time point of the monitoring horizon $1 \leq \tau \leq T_m$, and this estimate is compared against an estimate computed using the data in the training period $1 \leq t \leq m$. Conceptually, this would require using, in (3.13), $S_\tau - \tau S_m/m$, where $S_m = \sum_{t=1}^{m}(\psi_t + z_t)$ is defined with obvious notation. However, we know that S_m/m drifts to zero, and consequently we can directly use S_τ alone. Since we do not need to use a benchmark estimate based on the training sample, we can use the simple boundary function τ^η, which is a natural, non *ad-hoc*, threshold, based on the growth rate of partial sums of *i.i.d.* Gaussian sequences.

We define

$$\alpha_{T_m} = \sqrt{2\ln\ln T_m}, \quad \beta_{T_m} = 2\ln\ln T_m + \frac{1}{2}\ln\ln\ln T_m - \frac{1}{2}\ln\pi,$$

and let $\{W(u), 0 < u < \infty\}$ denote henceforth a standard Wiener process.

THEOREM 1. *We assume that Assumptions A.1-A.5 and 1-3 hold. Then, under the null, as* $\min\{p_1, p_2, m\} \to \infty$, *it holds that for almost all realisations of* $\{\mathbf{X}_t, 1 \leq t \leq T\}$

(i) *for all* $0 \leq \eta < 1/2$,

$$T_m^{\eta-1/2}\max_{1\leq\tau\leq T_m}\frac{|S_\tau|}{\tau^{\eta}} \xrightarrow{D^*} \sup_{0\leq u\leq 1}\frac{|W(u)|}{u^{\eta}}; \tag{3.18}$$

(ii) *for all* $-\infty < v < \infty$

$$P^*\left(\alpha_{T_m}\max_{1\leq\tau\leq T_m}\frac{|S_\tau|}{\tau^{1/2}} \leq v + \beta_{T_m}\right) = \exp\left(-\exp\left(-v\right)\right); \tag{3.19}$$

(iii) *for all* $\eta > 1/2$

$$r_{T_m}^{\eta-1/2}\max_{r_{T_m}\leq\tau\leq T_m}\frac{|S_\tau|}{\tau^{\eta}} \xrightarrow{D^*} \sup_{1\leq u<\infty}\frac{|W(u)|}{u^{\eta}}. \tag{3.20}$$

The theorem provides the asymptotics under the null for the functionals defined in (3.14)-(3.16). In the case of equation (3.18), the null of no break is rejected whenever

$$|S_\tau| > c_{\alpha,\eta}T_m^{-\eta+1/2}\tau^{\eta},$$

where $P\left(\sup_{0\leq u\leq 1} u^{-\eta}\left|W\left(u\right)\right|>c_{\alpha,\eta}\right)=\alpha$, thereby defining the changepoint estimate

$$\widehat{\tau}=\left\{\min_{1\leq\tau\leq T_m}\tau:\left|S_\tau\right|\geq c_{\alpha,\eta}T_m^{-\eta+1/2}\tau^\eta\right\}.$$

When using (3.15), equation (3.19) entails that rejection takes place, with a changepoint being identified, at $\widehat{\tau}=\left\{\min_{1\leq\tau\leq T_m}\tau:\left|S_\tau\right|\geq c_{\alpha,m}\tau^{1/2}\right\}$, where the asymptotic critical value $c_{\alpha,m}$ is defined as

$$c_{\alpha,m}=\frac{\beta_{Tm}-\ln\left(-\ln\left(1-\alpha\right)\right)}{\alpha_{Tm}}. \tag{3.21}$$

It is well known that convergence to the extreme value distribution can be slow; the simulations in Csörgő and Horváth (1997) suggest that the asymptotic critical values tends to overstate the correct ones: whilst our simulations show that results tend to be relatively good even when using asymptotic critical values, the fact that our randomisation scheme is additive, and based on z_τ being normally distributed, readily allows us to propose a refinement of critical values for the case $\eta=1/2$, which we discuss in Section B.1.5 in the Supplement. Finally, in the case of (3.16), rejection takes place at $\widehat{\tau}=\left\{\min_{1\leq\tau\leq T_m}\tau:\left|S_\tau\right|\geq c_{\alpha,\eta}r_{T_m}^{-\eta+1/2}\tau^\eta\right\}$, where $P\left(\sup_{1\leq u<\infty}u^{-\eta}\left|W\left(u\right)\right|>c_{\alpha,\eta}\right)=\alpha$. Note that, using the scale transformation of the Wiener process

$$\sup_{1\leq u<\infty}\frac{\left|W\left(u\right)\right|}{u^\eta}\overset{D}{=}\sup_{0\leq s\leq 1}\frac{\left|W\left(s\right)\right|}{s^{1-\eta}},$$

so that the critical values in Table 1 in Horváth et al. (2004) can be used here.

As a final remark, we would like to point out that the results in Theorem 1 are based on the strong approximation of the partial sums of z_τ (see Komlós et al., 1975, and Komlós et al., 1976). In our methodology, we assume that $z_\tau\sim i.i.d.N\left(0,1\right)$; however, in principle it would be possible (if sub-optimal) to use other distributions for $\{z_\tau,1\leq\tau\leq T_m\}$; in Section B.1.6 in the Supplement, we discuss alternative choices, explaining why a Gaussian distribution for $\{z_\tau,1\leq\tau\leq T_m\}$ yields the best results.

We now show that our procedures have power versus the alternatives considered in (3.2) and (3.4).

Assumption 4. *It holds that* $T_m-t^*=\Omega\left(m\right)$.

In Assumption 4, we assume that, if a break occurs, this does not happen too close to the end of the monitoring period.

Theorem 2. *We assume that Assumptions A.1-A.5 and 1-4 hold. Then, under (3.2) and (3.4), as* $\min\left\{p_1,p_2,m\right\}\to\infty$, *it holds that, for almost all realisations of* $\{\mathbf{X}_t,1\leq t\leq T\}$*:*

(i) *If, for* $0\leq\eta\leq 1/2$, *it holds that*

$$\frac{T_m^{1/2}}{\sqrt{\ln\ln T_m}}p_1^{r(1-\delta)}\to\infty, \tag{3.22}$$

then

$$T_m^{\eta-1/2}\max_{1\leq\tau\leq T_m}\tau^{-\eta}\left|S_\tau\right|\overset{P^*}{\to}\infty, \tag{3.23}$$

for all $0 \leq \eta < 1/2$*, and*

$$P^*\left(\alpha_{T_m} \max_{1\leq\tau\leq T_m} \tau^{-1/2} |S_\tau| > v + \beta_{T_m}\right) = 1, \tag{3.24}$$

for $\eta = 1/2$ *and all* $-\infty < v < \infty$.

(ii) *If, for* $\eta > 1/2$*, it holds that*

$$\left(\frac{r_{T_m}}{T_m}\right)^{\eta-1/2} \frac{T_m^{1/2}}{\sqrt{\ln\ln T_m}} p_1^{r(1-\delta)} \to \infty, \tag{3.25}$$

then

$$(r_{T_m})^{\eta-1/2} \max_{r_{T_m}\leq\tau\leq T_m} \tau^{-\eta} |S_\tau| \xrightarrow{P^*} \infty. \tag{3.26}$$

We conclude our discussion with some heuristic considerations on how the detection delay is affected by the combinations of m, p_1 and p_2. In particular, the size of the rolling window m is typically user-defined in the general context of sequential detection problems, and it is well-known to have an impact on the detection delay (see the analysis in Aue and Horváth, 2004, as an example). Towards the development of an adaptive selection rule for m (which we present in Section 4.1), we now discuss how m can affect the detection delay; details are in Section B.2.2 in the Supplement. Recall that, by (3.3) and (3.5), in the presence of a changepoint at time $t^* + 1$, the $(k_1 + 1)$-th eigenvalue calculated at $\tau \geq t^* + 1$ - denoted as $\widehat{\lambda}_{k_1+1,\tau}$ - will be proportional to $\frac{\tau-t^*}{m} p_1$. In the construction of our test statistics, we premultiply $\widehat{\lambda}_{k_1+1,\tau}$ by $p_1^{-\delta}$; this entails that, heuristically, a sufficient condition for changepoint detection is

$$\frac{p_1^{1-\delta}}{m} (\tau - t^*) \to \infty. \tag{3.27}$$

This result suggests that, when there are breaks that enlarge the factor space and increase the number of spiked eigenvalues, our methodology (which is based on the spiked eigenvalues) is able to exploit the fact that the row cross-sectional dimension p_1 passes to infinity: this ensures a short detection delay, and it is an example of the aforementioned blessing of dimensionality encountered in factor models. Low-dimensional methodologies, like the one proposed in Bai et al. (2022), would not have this advantage by construction.

3.1.2. *Monitoring schemes based on worst-case scenario.* In this section, we complement the theory developed in the previous section by proposing a different and completely novel monitoring scheme. We follow the same presentation layout as in the previous section: we begin by defining the monitoring scheme; we then study the limiting null distribution and investigate power against alternatives (see Theorem 3); and, finally, we discuss the delay in detecting changepoints.

Recall that, heuristically, the sequence $\{y_\tau, 1 \leq \tau \leq T_m\}$ is *i.i.d.* Gaussian under the null (conditional on the sample), whereas it diverges to positive infinity under the alternative. These features (independence and Gaussianity) arise directly from our randomisation scheme, and they allow to propose a monitoring scheme based on the "worst-case scenario", viz.

$$Z_{T_m} = \max_{1\leq\tau\leq T_m} y_\tau. \tag{3.28}$$

In order to study the asymptotics of Z_{T_m}, we define the norming sequences

$$a_{T_m} = \frac{b_{T_m}}{1 + b_{T_m}^2}, \; b_{T_m} = \sqrt{2 \ln T_m} - \frac{\ln \ln T_m + \ln (4\pi)}{2\sqrt{2 \ln T_m}},$$

which are proposed in Gasull et al. (2015).

THEOREM 3. *We assume that Assumptions A.1-A.4 and 1-4 hold. Then, under the null, as* $\min \{p_1, p_2, m\} \to \infty$, *it holds that*

$$\lim_{\min\{p_1,p_2,m\}\to\infty} P^* \left(\frac{Z_{T_m} - b_{T_m}}{a_{T_m}} \leq v \right) = \exp \left(- \exp \left(-v \right) \right), \tag{3.29}$$

for almost all realisations of $\{\mathbf{X}_t, 1 \leq t \leq T\}$ *and all* $-\infty < v < \infty$.

Under the alternatives (3.2) and (3.4), if it holds that

$$\frac{p_1^{r(1-\delta)}}{\sqrt{\ln T_m}} \to \infty, \tag{3.30}$$

as $\min \{p_1, m\} \to \infty$, *then it follows that*

$$\lim_{\min\{p_1,p_2,m\}\to\infty} P^* \left(\frac{Z_{T_m} - b_{T_m}}{a_{T_m}} \leq v \right) = 0, \tag{3.31}$$

for almost all realisations of $\{\mathbf{X}_t, 1 \leq t \leq T\}$ *and all* $-\infty < v < \infty$.

According to (3.29), rejection takes place, with a break being identified, at $\widehat{\tau} = \{\min \tau : y_\tau > c_{\alpha,2}\}$, where the asymptotic critical value $c_{\alpha,2}$ is given by

$$c_{\alpha,2} = b_{Tm} - a_{Tm} \ln \left(- \ln \left(1 - \alpha \right) \right). \tag{3.32}$$

It is well-known that convergence to the extreme value distribution can be slow (Hall, 1979). In our case, we found in our simulations that the asymptotic critical value in (3.32) works very well as far as our procedure is concerned; however, refinements are possible through the so-called "penultimate approximation", which we discuss in Section B.1.5 in the Supplement.

3.1.3. *Concluding remarks.* We conclude this section with some comments on the transformation ψ_τ defined in (3.10), which apply to both monitoring schemes in Sections 3.1.1 and 3.1.2 (see also Section B.2.3 in the Supplement for further details). As mentioned above, the choice of r in (3.10) directly affects the magnitude of ψ_τ. Under the null hypothesis, ψ_τ drifts to zero faster as r gets larger (and vice versa); this is desirable, since - under the null - the presence of ψ_τ adds a non-centrality parameter to the weak limit of the test statistic. Conversely, under the alternative ψ_τ passes to infinity faster as r gets larger (and vice versa); this also is desirable, since such divergence triggers the detection of a changepoint. Hence, in our randomisation scheme, there is no trade-off between size and power: as far as the choice of r is concerned, from a practical point of view r should be chosen as large as possible. Thus, the choice of r is not a matter of tuning, but rather an entirely data-driven choice which arises from the largest moment of the common factors $\mathbf{F}_t$ and of the innovations $e_{ij,t}$. Assumptions A.1 and A.3 indicate that, if, for example, the 8-th moments of $\mathbf{F}_t$ and $e_{ij,t}$ exist at most, then $r = 2$, and so on.

4. Guidelines for implementation and simulations.

4.1. *Implementation guidelines.* As discussed above, the practical implementation of our procedure requires some specifications: *(a)* the "nuisance parameter" r, i.e. the maximal moment admitted by the data; *(b)* the tuning parameter ε in (3.9); and *(c)* as is typically the case in any sequential changepoint problem, it requires deciding the size of the training window m.

We begin by discussing the choice of r, and therefore the construction of ψ_τ in (3.10). Our remarks at the end of Section 3.1.1 indicate that r should be as large as possible, and that this is beneficial under both the null and the alternative. Usually, the maximal moment r is not known. In practical applications, r can be determined in at least two ways. A "bottom-up" way is based on estimating the tail index of the $X_{ij,t}$s (e.g., via one of the estimators discussed in De Haan and Ferreira, 2007); a "top-down" way is based on choosing a priori a possible value of r and applying a test for $H_0 : E\left|X_{ij,t}\right|^{4r} = \infty$ such as the one by Trapani (2016) (see also Fedotenkov, 2013), and, upon rejecting H_0, using the chosen value of r. Alternatively and/or in addition, data pre-processing (e.g. windsorisation or outlier removal) can be implemented. In Section B.2.3 of the Supplement, we assess the sensitivity of our methodology to the value of r, also adding some guidelines on how to implement the "top-down" approach proposed above.

After determining r, we turn to considering the choice of the tuning parameter δ, and more specifically of ε in (3.9) - the rest of (3.9), i.e. β and whether this is greater or smaller than $1/2$, is completely data-driven and hence not a matter of tuning. Indeed, ε in (3.9) is the only tuning parameter which is specifically required by our procedure. The rationale underpinning the choice of ε in (3.9) is described in Section 3: multiplying $\widehat{\lambda}_{k_1+1,\tau}$ by $p_1^{-\delta}$ serves the purpose of making $p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}$ drift to zero under the null (essentially, by smoothing away the estimation error); at the same time, under the alternative, $p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}$ should diverge to positive infinity. Hence, the choice of ε reflects, the trade-off between size and power. In this respect, m and ε interact with each other: if both are large, convergence under the null will take place at a faster rate ensuring better size control (but at the price of slower detection of changepoints), and vice versa if both are small, detection delays will be shorter (but at the price of potential size distortion). On account of this, from a practical point of view our recommendation is to firstly fix ε in (3.9), and then select m. In Section B.3.2 in the Supplement, we assess via simulation the sensitivity of our results to ε in (3.9): the results show that even a small value of ε suffices in making $p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}$ drift to zero under the null at a satisfactory rate. Hence, our recommendation is to set $\varepsilon = 0.05$, but as our simulations show, results are virtually unaffected when ε is increased or reduced.

We conclude our flowchart by providing an adaptive rule to choose an upper bound for m. Our asymptotics requires $m \to \infty$; whilst no lower bound is explicitly required, upon inspecting our proofs it emerges that - given δ - the larger m, the better our asymptotic approximations. On the other hand, equation (3.27) suggests that as m gets larger, this could worsen the detection delay under the alternative, thus indicating that the choice of m reflects the trade-off between power and size. Indeed, as also mentioned above, the specification of m is required in every applications of online changepoint detection; as Chu et al. (1996) put it, "it seems attractive [...] to make the choice of m part of the monitoring design scheme" (p. 1058). A possible approach can be based on having a rule which, after ensuring convergence under the null to the appropriate limiting distributions, chooses m so as to minimise the detection delay - this is based on combining (3.27) with the bound in (3.12). The upper bound for m, which ensures a short detection delay satisfies

$$m = O\left(p_2^{r/(r+2)-\epsilon}\right), \tag{4.1}$$

for any $\epsilon > 0$. In Section B.2.2 in the Supplement, we report a derivation of (4.1). Equation (4.1) illustrates the relationship between m and the maximal moment r: the larger the latter, the closer m can be to p_2. We conclude by pointing out that (4.1) ensures a short detection delay, but it may not always be appropriate in practical applications: when p_2 is small, (4.1) may yield a very small m which, in practice, may not suffice to ensure that the limiting null distribution is a good approximation. In such cases, a larger, sub-optimal value of m may be needed, which, based on (3.27), would yield a larger detection delay. Thus, in order to complement the theoretical rule in (4.1), in Section B.3.3 in the Supplement we also report a comprehensive set of simulations to assess the sensitivity of our methodology to the choice of m. An R package "OLCPM" implementing the procedure is available on CRAN.

4.2. *Monte Carlo analysis.* We assess the performance of the proposed procedures on testing and locating changepoints through synthetic data. Throughout the section, the data generating process is similar to the one used in Yu et al. (2022) and He et al. (2023). Specifically, under the null hypothesis H_0 without change point, we set the row/column factor numbers $k_1 = k_2 = 3$. The entries of $\mathbf{R}$ and $\mathbf{C}$ are independently sampled from uniform distribution $\mathcal{U}(-1, 1)$, while

$$\text{Vec}(\mathbf{F}_t) = \phi \text{Vec}(\mathbf{F}_{t-1}) + \sqrt{1-\phi^2}\text{Vec}(\epsilon_t), \quad \epsilon_t \sim i.i.d.\ \mathcal{N}(0, I_{k_1 \times k_2}), \tag{4.2}$$

$$\text{Vec}(\mathbf{E}_t) = \psi \text{Vec}(\mathbf{E}_{t-1}) + \sqrt{1-\psi^2}\text{Vec}(\mathbf{U}_t), \quad \mathbf{U}_t \sim i.i.d.\ \mathcal{N}(0, \mathbf{V}_E \otimes \mathbf{U}_E), \tag{4.3}$$

where $\mathbf{U}_t$ is from a matrix-normal distribution, i.e., $\text{Vec}(\mathbf{U}_t) \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{V}_E \otimes \mathbf{U}_E)$, and $\mathbf{V}_E$ and $\mathbf{U}_E$ are matrices with ones on the diagonal, and the off-diagonal entries are a/p_1 and a/p_2, respectively. The parameter a controls cross-sectional dependence, with larger a leading to stronger cross-dependence; we have used $a = 2$ in our simulations, similarly to He et al. (2023). In (4.2)-(4.3), the parameters ϕ and ψ control both the temporal and cross-sectional correlations of $\mathbf{X}_t$. In the simulation study, we let $\psi = \phi = 0.1$.

Throughout all experiments reported here, we have used $p_1 \in \{50, 80, 100\}$, $p_2 \in \{20, 50, 80\}$, and $m = p_2$ - as far as the latter choice is concerned, we note that it is related to (4.1) upon observing that all moments exist, and that, when p_2 is small, m cannot be too small or size control may be affected. The monitoring procedures are based on the 4-th largest eigenvalue of the rolling column-column sample covariance matrix. When calculating the projection matrix $\widetilde{\mathbf{C}}$, we always use $k_2 = k_{\max} = 8$ instead of estimating k_2 - this is suggested by the simulations in He et al. (2023), but we point out that preliminary, unreported experiments using $\widetilde{k}_2$ discussed in Section B.2.1 resulted in essentially the same results. In order to calculate the sequences ψ_τ, we let $\varepsilon = 0.05$ in (3.9), and use a notional value of $r = 8$ - increasing this, as would be allowed by having Gaussian data, does not alter results in any significant way. All results have been obtained using $1,000$ replications. We compare the performance of our methodology, based on using the projection-based estimator of Yu et al. (2022), versus a "matrix flattening" approach based on the spectrum of $\mathbf{M}_r$ defined in (2.2); in Section B.3.7 in the Supplement, we also report a brief set of experiments using the methodology of Barigozzi and Trapani (2020), applied to $\text{Vec}(\mathbf{X}_t)$; in this case, results are very poor, with the procedure having no power at all versus any changepoint.

Table 1 reports the empirical rejection frequencies under the null, for various values of η when using procedures based on partial sums. It is worth pointing out that, in the context of online changepoint detection, size control is different than in the standard Neyman-Pearson testing context; as Horváth et al. (2007) put it, "the goal is to keep the probability of false rejection below α rather than to make it close to α". Considering the methodology based on the projection-based estimator, all the empirical sizes are controlled (with very few exceptions, e.g. the case of small

TABLE 1

Empirical sizes *(×100) under H_0 over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\epsilon = 0.05$. "PS", partial-sum; "WC", worst case.*

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
200	20	80	4.5	1.4	2.8	3.3	35.7	20.3	8.4	7.2
200	30	80	5.0	1.6	3.2	4.0	41.9	27.5	8.7	8.8
200	40	80	4.3	1.7	2.9	4.3	43.3	24.8	8.6	8.4
200	80	20	4.0	1.8	3.1	5.1	4.0	2.3	2.9	3.1
200	80	30	5.0	2.6	3.3	4.9	3.4	1.7	3.2	4.3
200	80	40	4.4	1.9	2.6	4.2	4.0	2.2	4.2	4.2
200	20	20	11.6	7.0	4.2	5.5	62.6	50.1	29.1	19.2
200	30	30	5.8	2.4	4.1	3.3	63.0	49.2	23.8	14.4
200	40	40	4.3	2.0	3.4	3.5	61.2	42.8	19.8	10.7
80	40	40	3.9	1.7	2.5	3.2	27.9	15.3	7.3	7.3
120	40	40	5.4	1.8	2.9	3.7	40.3	25.5	11.1	10.1
160	40	40	4.8	1.7	2.1	4.3	52.6	37.2	14.4	9.9
200	40	40	3.4	1.5	2.5	3.2	62.2	46.5	18.0	10.8
240	40	40	3.8	1.8	2.5	3.6	68.8	50.7	18.9	11.2
280	40	40	3.5	1.4	3.1	3.6	72.4	55.7	21.3	10.8
320	40	40	3.7	1.8	2.7	5.7	76.3	60.5	19.5	11.5
360	40	40	3.9	1.9	1.8	4.7	80.5	65.1	22.3	13.1
400	40	40	3.9	1.9	2.5	4.0	80.7	63.9	20.0	11.6

p_1, p_2 and m using $\eta = 0.25$) at the given significance level, even in the case of small samples; in unreported simulations, we have also tried $\alpha = 0.1$, with the same results. The empirical rejection frequencies are close to their theoretical significance levels when using the partial-sum method with $\eta < 0.5$, and when using the worst-case scenario method described in Section 3.1.2; the latter finding is interesting, since convergence to the extreme value distribution can be slow. When $\eta = 0.5$, the empirical sizes are usually smaller than the significance levels α, as also noted in Gombay and Horváth (1996). Similarly, when $\eta > 0.5$, the empirical sizes also tends to be smaller than the theoretical level α, mainly because in this case (similarly to the Darling-Erdős case where $\eta = 0.5$) the effective sample size that determines the asymptotic distribution is smaller than T_m; similar results have also been observed in Horvath and Trapani (2022) in the context of in-sample changepoint detection. In the Supplement (Section B.1.5), we also propose refinements to the extreme value theory discussed here, and report a small set of experiments to assess them. Finally, recall that by way of comparison we have also reported results obtained using the spectrum of $\mathbf{M}_r$, i.e. not using the projection-based estimator. In this case, size control is lost in the vast majority of cases, thereby strongly suggesting that a methodology specifically designed for matrix-values series is required, and "flattening" approaches based on treating the observations $\{\mathbf{X}_t, 1 \leq t \leq T\}$ as a vector may not work.

We now study the power of our procedure to detect changepoints. As a first alternative, we consider (3.2) without $\mathbf{R}_0$, that is we consider the case where all loadings change after a changepoint located at $t^* = 0.5T$, viz.

$$\mathbf{X}_t = \begin{cases} \mathbf{R}_1 \mathbf{F}_{1,t} \mathbf{C}' + \mathbf{E}_t, & 1 \leq t \leq t^*, \\ \mathbf{R}_2 \mathbf{F}_{1,t} \mathbf{C}' + \mathbf{E}_t, & t^* + 1 \leq t \leq T, \end{cases}$$

where $\mathbf{R}_2$ is generated after time point t^*+1, with a fraction γ of its entries sampled from *i.i.d.* uniform distribution $\mathcal{U}(-1,1)$ (the remaining entries are the same as in $\mathbf{R}_1$) - hence, γ controls the magnitude of the break, and we study the sensitivity of our results to this parameter in Section B.3.4 in the Supplement. In Table 2, we report the empirical rejection frequencies; these are almost always close to 1, thus suggesting that our procedures will, eventually, find evidence of a changepoint if present. The power increases with p_1, as can be expected from the theory, and also as p_2 increases (this is due to the fact that, *ceteris paribus*, a larger p_2 entails that δ is smaller); indeed, there are some cases where the power is below 50%, but this immediately picks up as the sample sizes increase. Detection delays are reported in Table 3: based on those results, we conclude that our procedures tend to have short detection delays in all scenarios considered, even when the cross-sectional dimensions p_1, p_2 are small, offering accurate and early detection. Interestingly, results do not seem to be affected by T_m. We note that results look very good when using the spectrum of $\mathbf{M}_r$, but of course this needs to be read in conjunction with the loss of size control discussed above.

TABLE 2

Empirical powers *(×100) under H_{A1} over 1000 replications, with significance level $\alpha = 0.05$, $\gamma = 0.30$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\varepsilon = 0.05$. "PS", partial-sum; "WC", worst case.*

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
T	p_1	p_2	$\eta=0.25$	$\eta=0.5$	$\eta=0.75$		$\eta=0.25$	$\eta=0.5$	$\eta=0.75$	
200	20	80	80.8	77.8	74.0	74.5	90.0	86.2	74.1	67.9
200	30	80	98.5	97.9	96.8	97.0	99.2	98.4	95.9	94.1
200	40	80	99.9	99.8	99.5	99.6	100.0	99.9	99.3	98.9
200	80	20	30.8	25.5	15.9	38.8	8.8	4.6	4.3	9.7
200	80	30	96.0	94.1	91.0	96.0	64.1	59.6	49.2	66.8
200	80	40	99.9	99.9	99.8	99.9	97.3	96.6	94.4	96.3
200	20	20	64.7	57.7	45.6	63.8	85.9	79.3	58.7	64.3
200	30	30	95.1	93.8	91.3	94.7	98.9	97.5	92.1	93.1
200	40	40	99.8	99.8	99.2	99.8	99.9	99.8	98.8	98.9
80	40	40	99.8	99.8	99.5	99.5	99.9	99.9	99.3	99.2
120	40	40	99.8	99.8	99.6	99.5	99.9	99.8	99.5	99.1
160	40	40	99.5	99.4	99.1	99.2	99.8	99.5	98.6	98.3
200	40	40	99.8	99.7	99.4	99.8	100.0	100.0	99.4	99.3
240	40	40	100.0	99.9	99.8	99.9	99.9	99.8	98.9	99.5
280	40	40	99.9	99.8	99.2	99.8	99.9	99.9	98.5	99.2
320	40	40	99.6	99.4	98.6	99.7	99.9	99.9	97.7	98.8
360	40	40	99.7	99.6	98.7	99.8	100.0	99.9	98.5	99.0
400	40	40	99.4	99.2	98.4	99.4	100.0	99.8	97.8	98.8

Essentially the same results are found when considering the alternative (3.4) - i.e. the case where the number of common factors increases after the changepoint. We generate data according to

$$\mathbf{X}_t = \begin{cases} \mathbf{R}\mathbf{F}_{1,t}\mathbf{C}' + \mathbf{E}_t, & 1 \le t \le t^*, \\ \widetilde{\mathbf{R}}\mathbf{F}_{2,t}\mathbf{C}' + \mathbf{E}_t, & t^*+1 \le t \le T, \end{cases}$$

where $\mathbf{F}'_{2,t} = \left[\mathbf{F}'_{1,t} | \widetilde{\mathbf{F}}'_{1,t}\right]$ and $\widetilde{\mathbf{R}} = [\mathbf{R}|\mathbf{R}_2]$ and $\mathbf{R}_2$ is a $p_1 \times 1$ vector with a fraction γ of its entries sampled from *i.i.d.* uniform distribution $\mathcal{U}(-1,1)$ (the other entries are set equal to zero), and $\widetilde{\mathbf{F}}_{1,t}$ are the additional $1 \times k_2$ factors, whose entries are generated as *i.i.d.*$N(0,1)$. Therefore, after the

TABLE 3

Empirical powers *(×100) under H_{A2} over 1000 replications, with significance level $\alpha = 0.05$, $\gamma = 0.30$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\varepsilon = 0.05$. "PS", partial-sum; "WC", worst case.*

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
200	20	80	66.1	60.5	54.6	59.1	76.1	68.9	53.4	53.1
200	30	80	95.8	94.8	92.8	94.1	97.5	95.2	91.0	91.1
200	40	80	99.7	99.7	99.4	99.5	99.8	99.6	98.9	99.0
200	80	20	84.3	77.2	63.9	79.3	39.7	30.7	16.1	31.6
200	80	30	100.0	99.8	99.2	99.5	93.7	91.5	83.7	87.7
200	80	40	100.0	100.0	100.0	100.0	99.8	99.7	99.6	99.7
200	20	20	84.6	81.3	71.7	78.2	94.0	90.8	79.0	77.6
200	30	30	97.8	97.3	96.0	97.0	99.4	98.9	96.7	95.5
200	40	40	100.0	100.0	99.7	99.9	99.9	99.9	99.7	99.6
80	40	40	98.8	98.5	97.4	98.6	98.7	98.2	96.2	97.5
120	40	40	99.6	99.3	99.2	99.3	99.8	99.7	98.5	98.2
160	40	40	99.8	99.6	99.3	99.4	99.9	99.7	99.2	99.1
200	40	40	99.9	99.8	99.4	99.7	99.9	99.8	99.5	99.3
240	40	40	99.7	99.5	99.1	99.3	99.9	99.9	99.4	99.3
280	40	40	100.0	100.0	99.6	99.8	100.0	100.0	99.3	98.8
320	40	40	99.9	99.9	99.8	99.7	100.0	99.9	99.9	99.7
360	40	40	99.9	99.9	99.7	99.7	100.0	99.9	99.8	99.4
400	40	40	99.9	99.8	99.8	99.8	100.0	100.0	99.8	99.7

changepoint, the number of row factors grows to 4, and we are monitoring according to the 4-th eigenvalue of the rolling column-column sample covariance matrix. The other parameters are set exactly the same as those introduced above. Results are reported in Tables 4 and 5. We note that delays appear to be slightly larger than those reported for alternative hypothesis (3.2) in Table 3. However, the monitoring procedures can still provide accurate and timely detection when change occurs under (3.4), even in small samples cases. Note that the impact of the sample sizes m, p_1 and p_2 is exactly the same as under (3.2).

We conclude by pointing out that, in the Supplement, we report further simulations under different scenarios (Section B.3): we consider sensitivity to k_2, δ and m in Sections B.3.1, B.3.2 and B.3.3 respectively. We assess the impact on size and/or power of the size of change (Section B.3.4), of the presence of changes in $\mathbf{C}$ (Section B.3.5), and of smooth transitioning breaks (Section B.3.6). Finally, we report the outcome of sequential monitoring using the method developed by Barigozzi and Trapani (2022) applied to the vectorised data (Section B.3.7), and we report further statistics on detection delays and computational times (Section B.3.8).

5. Empirical application. In this section, we illustrate the usefulness of our sequential monitoring schemes through an application to financial data. We use the Fama and French 10×10 series, which has been considered in several applications in the context of matrix factor models (see e.g. Wang et al., 2019; Yu et al., 2022; and Liu and Chen, 2022).[2] The dataset comprises monthly market-adjusted return series, with portfolios being the intersections of 10 portfolios formed by size (market equity, ME) and 10 portfolios formed by the ratio of book equity to market equity

[2]In Section B.3.9 in the Supplement, we also consider an application to macro data.

TABLE 4

Empirical delays *(interquantiles) under H_{A1} over 1000 replications, with significance level $\alpha = 0.05$, $\gamma = 0.30$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\varepsilon = 0.05$. "PS", partial-sum; "WC", worst case.*

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
200	20	80	16,21,28	16,20,27	17,22,28	13,18,24	17,22,31	16,22,31	18,23,33	15,20,28
200	30	80	12,15,20	11,15,20	12,16,21	10,13,18	13,17,22	12,17,22	14,19,24	12,16,21
200	40	80	9,11,14	9,11,14	10,12,16	7,9,12	10,13,16	9,12,16	11,14,18	9,11,15
200	80	20	7,9,12	7,9,11	6,9,11	5,7,8	8,15,59	-20,7,10	-73,-70,9	4,7,10
200	80	30	6,8,10	6,8,10	7,9,12	5,6,8	9,11,14	9,11,14	9,12,15	6,8,11
200	80	40	6,7,9	6,7,9	7,8,10	4,6,7	8,10,13	8,11,13	10,12,15	6,8,10
200	20	20	5,7,11	5,7,10	6,8,11	4,5,7	-28,5,10	-42,4,10	-62,6,11	3,5,8
200	30	30	5,7,10	5,7,10	6,9,11	4,6,7	-4,6,9	-6,6,9	5,9,12	4,6,9
200	40	40	5,7,9	5,7,9	6,8,10	4,5,7	3,6,9	4,6,9	6,9,11	4,6,8
80	40	40	4,6,8	4,6,8	5,7,8	4,5,7	5,7,9	5,6,8	5,7,9	5,6,8
120	40	40	4,6,8	4,6,8	6,7,10	4,5,7	5,7,9	5,7,9	6,8,11	4,6,8
160	40	40	5,7,9	5,7,9	6,8,10	4,5,7	4,6,9	4,7,9	6,9,11	4,6,8
200	40	40	5,7,9	5,7,9	6,8,10	4,5,7	3,6,9	4,6,9	6,8,11	4,6,8
240	40	40	5,7,9	5,7,9	6,8,10	4,5,7	-4,6,9	-1,6,9	6,8,11	4,6,8
280	40	40	5,7,9	5,7,9	6,8,11	4,5,7	-24,6,8	-34,6,9	6,9,12	4,6,8
320	40	40	5,7,9	5,7,9	6,8,11	4,5,7	-29,5,8	-33,6,9	6,9,12	4,6,8
360	40	40	5,7,9	5,7,9	6,8,11	4,5,7	-46,4,8	-61,5,9	6,9,12	4,6,8
400	40	40	5,7,9	5,7,9	7,9,11	4,5,7	-56,4,8	-69,5,8	6,9,12	4,6,8

TABLE 5

Empirical delays *(interquantiles) under H_{A2} over 1000 replications, with significance level $\alpha = 0.05$, $\gamma = 0.30$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\varepsilon = 0.05$. "PS", partial-sum; "WC", worst case.*

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
200	20	80	51,64,76	52,64,77	54,67,80	47,60,73	42,56,72	44,60,76	52,68,82	51,64,77
200	30	80	36,45,56	37,46,57	40,50,61	34,44,55	33,43,56	34,46,59	41,53,66	38,49,62
200	40	80	29,35,43	29,36,43	31,39,47	27,33,41	28,35,44	29,37,46	34,42,52	31,38,48
200	80	20	24,37,57	24,38,58	26,41,63	19,28,49	35,56,78	39,58,83	27,49,76	24,43,72
200	80	30	20,24,29	20,24,29	22,27,32	17,21,26	27,32,43	27,33,45	29,37,50	23,28,36
200	80	40	18,22,26	18,22,26	20,24,29	16,19,23	25,29,35	26,30,36	28,33,39	22,27,32
200	20	20	18,24,40	18,25,43	21,29,47	16,21,36	-27,15,28	-37,17,32	-35,23,42	14,21,39
200	30	30	17,22,27	17,22,28	19,24,30	14,19,25	-6,17,24	-6,19,26	15,24,31	15,21,27
200	40	40	16,20,25	17,21,25	18,23,28	14,18,22	10,18,23	12,19,25	18,24,29	16,20,25
80	40	40	15,19,23	15,19,24	16,20,25	14,18,22	16,20,24	16,20,25	17,22,27	16,20,25
120	40	40	16,20,24	16,20,25	17,22,27	14,18,22	15,19,25	15,20,26	18,23,29	16,20,25
160	40	40	16,20,24	16,20,25	18,22,27	14,18,22	13,19,24	14,20,25	18,24,30	16,20,25
200	40	40	17,20,25	17,21,25	18,23,27	14,18,22	5,17,24	8,19,25	17,23,29	16,20,25
240	40	40	17,21,27	17,21,27	19,24,29	14,18,23	-1,17,24	0,19,26	18,25,31	16,21,27
280	40	40	17,21,26	17,21,27	19,24,30	14,18,23	-16,16,24	-21,18,25	18,24,31	16,21,26
320	40	40	17,21,27	18,22,28	20,24,31	14,18,24	-32,15,23	-35,17,25	18,25,32	16,21,27
360	40	40	17,22,27	18,22,28	20,25,30	14,18,23	-47,13,23	-62,16,24	18,25,32	16,21,26
400	40	40	18,22,27	18,22,27	21,25,30	14,18,23	-65,9,23	-80,15,24	18,26,32	16,21,27

(BE/ME), which leads to 10×10 matrix-variate observations.[3] We collect portfolio series from January 1964 to December 2020, totalling 684 months. Missing values (missing rate is 0.25%) are inputted by linear interpolation. Subsequently, we follow the preprocessing procedures in Wang et al. (2019) and Yu et al. (2022), subtracting the monthly market excess returns and standardising each series.

We implement our procedure as follows. We use both ψ_τ and $\widetilde{\psi}_\tau = 1/\psi_\tau$ (see Section B.1.1 in the Supplement) to detect breaks involving an increase in the number of factors/changes in the loadings and vanishing factors respectively. We use both the projection-based estimator, and the non-projected methodology. We consider both methodologies based on the fluctuations of partial sums, and the worst-case scenario statistic; the results in Table 1 suggest that the latter leads to better size-control when p_1 and p_2 are small. All our tests are carried out at $\alpha = 0.05$ nominal level.

In order to ensure that a sufficient number of moments exists, we remove outliers (defined as observations deviating from the median more than 3 times interquartile range). By way of double-checking, we also applied the test by Trapani (2016) for $H_0 : E\left|X_{ij,t}\right|^{32} = \infty$,[4] rejecting the null; this result informs the choice, in our procedure, of $r = 8$ in (3.10), as also used in the Monte Carlo experiments. As far as the size of the training sample, m, is concerned, we note that, in this dataset, p_2 is quite small, which suggests that the rule designed in (4.1) may not ensure an adequate size control; therefore, we use a training period of size $m = 24$ (i.e., 2 years) for all the monitoring schemes. This is not the optimal choice as far as detection delay reduction is concerned, but our simulations suggest that it should yield good size control. Finally, our simulations in Table 1 indicate that the proposed tests are more likely to lose control of size whenever p_1 and p_2 are small. Hence, in equation (3.9), we use a slightly larger $\varepsilon = 0.25$ to calculate δ when testing for the increase of factor number or the change of loading space. Conversely, in Section B.1.1 we note that size control is easier when testing for vanishing factors, and we therefore use the (recommended) $\varepsilon = 0.05$ in that case. Finally, by way of robustness check, we carry out our monitoring procedure 100 times, and declare a change point only when over 75% of replications reject the corresponding null hypothesis; we estimate the breakdate as the median of detected values across the 100 replications. Whenever a change point is detected, we record the location τ_j^* and the type of change. Then, we drop all the series before τ_j^* and restart the monitor procedure, until the length of the remaining series is smaller than m or no change points are reported any more.

The first step of our analysis is to determine the numbers of row and column factors k_1, k_2 in the training period. Both when using the iterative algorithm in Yu et al. (2022), and the randomized testing procedure in He et al. (2023), we find that $k_1 = k_2 = 1$. Indeed, Figure 1 illustrates the four leading eigenvalues of the rolling column-column sample covariance matrices with and without projection. As shown in the figures, the first largest eigenvalue is always much larger than the remaining ones.

The full-blown set of results is in Table B.18 in the Supplement. At most four change points are found for the row factors, and the reported locations by different testing approaches are quite close. The locations of the change points are plotted in Figure 1, which coincide with the variation of the second largest rolling eigenvalue, although with slight delays. As far as results obtained with the projection-based method are concerned, the gap between the first two change points is short and

[3] Data have been downloaded from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[4] See Section B.2.3 in the Supplement

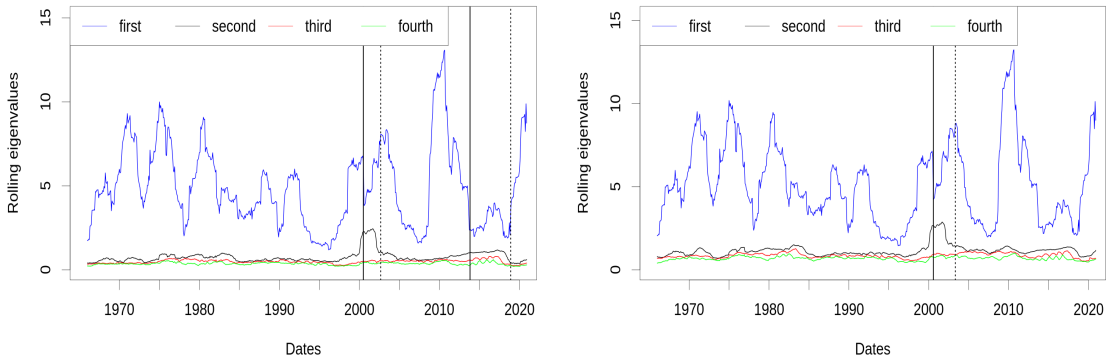

FIG 1. *Eigenvalues of the rolling column-column sample covariance matrices with (left) and without (right) projection in the monitoring, for the Fama-French 100 portfolios. The real vertical lines are the locations when loading space changes or the number of factors increases, while the dashed vertical lines are the locations when the number of factors decreases, respectively.*

roughly equal to m; the estimated breakdates (around August 2000 and October 2002) are highly suggestive, being around the dot.com bubble. On the other hand, the gap between the third and fourth change points are much longer, suggesting that probably a new row factor occurs and then vanishes during the period between November 2013 and December 2018; the 2018 date is harder to interpret, but we note that our methodology detects vanishing changepoints with a possibly lengthy delay. Conversely, the worst-case statistic without projection only finds the first two change points. Finally, we point out that we have also tried vectorising the matrix observations first and then apply the monitoring procedure proposed by Barigozzi and Trapani (2022). However, it turns out that this naive approach fails to find any change point in this example.
In the Supplement, we investigate the effect of the tuning parameters r, m, and ε (Table B.17), and we also investigate the presence of changepoints in the column factor structure (Table B.18).

6. Conclusions. In this paper, we have proposed several schemes for the online detection of changes in the latent factor structures of a two-way, matrix factor model. Our approach is based on noting that many instances of changepoint can be represented as a change in the dimension of the factor spaces. Hence, in order to detect changes, we use the eigenvalues of the projected second moment matrices, exploiting the fact that when a new factor emerges, the corresponding eigenvalue will be spiked. Crucially, since we do not know the limiting distribution of the estimated eigenvalues, we rely only on rates, and randomise the sequence of the estimated eigenvalue by perturbing it with an *i.i.d.*, standard normal sequence. Thence, we are able to propose two families of sequential procedures, one more "classical" and based on the fluctuations of partial sums, and another one, completely novel, based on the extreme value behaviour of the perturbed sequence of estimated eigenvalues. Our approach has several distinctive advantages: it is easy to implement, it requires much less tuning than competing approaches (see e.g. Barigozzi and Trapani, 2020), and it works very well in simulations, offering good size control and fast changepoint detection. Further, to the best of our knowledge, this is the first paper which designed an *ad-hoc* procedure for the case of matrix-valued time series, exploiting on the one hand the dimensionality of the series, and on the other hand the matrix structure of the data. Monitoring the eigenvalues of the second moment matrix of the data has several advantages, including exploiting the cross-sectional dimensions p_1

and p_2, and offering tests that are robust to changes in the second moment of the common factors.

For the sake of a concise discussion, we report only the main results in this manuscript. However, several interesting extensions can be considered. In Section B.1 in the Supplement we study: *(i)* the case of vanishing factors; *(ii)* the cases $k_1 = 0$ and/or $k_2 = 0$; *(iii)* the case of changepoint detection in the presence of weak factors; *(iv)* the robustness of our procedure in the presence of changes in the column factor loadings matrix $\mathbf{C}$; and *(v)* other results concerning the application of our methodology. Further, the case where the numbers of common factors k_1 and k_2 is not observed is studied in Section B.2.1.

As a by-product, our methodology for the online detection of changepoints could be generalised to other contexts; indeed, and immediate generalisation is to tensor-valued time series, which is a high priority on our agenda. From a methodological point of view, our procedure relies, essentially, on constructing an (asymptotically, and conditionally on the sample) *i.i.d.*, standard normal sequence under the null, and thereafter using such a sequence to construct detectors based e.g. on the CUSUM process. In principle, other detectors could also be constructed, such as MOSUM-based ones (see e.g. Eichinger and Kirch, 2018, and Dette and Gösmann, 2020), and we leave this (exciting) possibility for future work. Similarly, although in this paper we have focused on online, sequential detection, our procedure can also be extended to the case of offline, ex-post detection, offering an alternative to Bai et al. (2022) - indeed, our methodology is based on constructing an *i.i.d.* sequence, and detecting breaks in its mean. Hence, all the existing theory for the detection of a change in mean (see e.g. Csörgő and Horváth, 1997) can be readily employed.

References.

Aue, A., Horváth, L., 2004. Delay time in sequential detection of change. Statistics & Probability Letters 67, 221–231.

Bai, J., Duan, J., Han, X., 2022. Likelihood ratio test for structural changes in factor models. arXiv preprint arXiv:2206.08052 .

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Baltagi, B.H., Kao, C., Wang, F., 2017. Identification and estimation of a large factor model with structural instability. Journal of Econometrics 197, 87–100.

Baltagi, B.H., Kao, C., Wang, F., 2021. Estimating and testing high dimensional factor models with multiple structural changes. Journal of Econometrics 220, 349–365.

Bandi, F., Corradi, V., 2014. Nonparametric nonstationarity tests. Econometric Theory 30, 127–149.

Barigozzi, M., Cavaliere, G., Trapani, L., 2022. Inference in heavy-tailed non-stationary multivariate time series. Journal of the American Statistical Association *forthcoming* , 1–51.

Barigozzi, M., Cho, H., Fryzlewicz, P., 2018. Simultaneous multiple change-point and factor analysis for high-dimensional time series. Journal of Econometrics 206, 187–225.

Barigozzi, M., Trapani, L., 2020. Sequential testing for structural stability in approximate factor models. Stochastic Processes and their Applications 130, 5149–5187.

Barigozzi, M., Trapani, L., 2022. Testing for common trends in nonstationary large datasets. Journal of Business & Economic Statistics 40, 1107–1122.

Breitung, J., Eickmeier, S., 2011. Testing for structural breaks in dynamic factor models. Journal of Econometrics 163, 71–84.

Chen, E.Y., Chen, R., 2020. Modeling dynamic transport network with matrix factor models: with an application to international trade flow arXiv:1901.00769.

Chen, E.Y., Fan, J., 2021. Statistical inference for high-dimensional matrix-variate factor models. Journal of the American Statistical Association , 1–18.

Chen, E.Y., Tsay, R.S., Chen, R., 2020a. Constrained factor models for high-dimensional matrix-variate time series. Journal of the American Statistical Association 115, 775–793.

Chen, E.Y., Xia, D., Cai, C., Fan, J., 2020b. Semiparametric tensor factor analysis by iteratively projected SVD. arXiv preprint arXiv:2007.02404 .

Chen, L., Dolado, J.J., Gonzalo, J., 2014. Detecting big structural breaks in large factor models. Journal of Econometrics 180, 30–48.

Chen, R., Yang, D., Zhang, C.H., 2022. Factor models for high-dimensional tensor time series. Journal of the American Statistical Association 117, 94–116.

Chen, X., Yang, D., Xu, Y., Xia, Y., Wang, D., Shen, H., 2021a. Testing and support recovery of correlation structures for matrix-valued observations with an application to stock market data. Journal of Econometrics .

Chen, Y., Wang, T., Samworth, R.J., 2020c. High-dimensional, multiscale online changepoint detection. arXiv preprint arXiv:2003.03668 .

Chen, Y., Wang, T., Samworth, R.J., 2021b. Inference in high-dimensional online changepoint detection. arXiv preprint arXiv:2111.01640 .

Cheng, X., Liao, Z., Schorfheide, F., 2016. Shrinkage estimation of high-dimensional factor models with structural instabilities. The Review of Economic Studies 83, 1511–1543.

Chu, C.S.J., Stinchcombe, M., White, H., 1996. Monitoring structural change. Econometrica 64, 1045–1066.

Corradi, V., Swanson, N.R., 2006. The effects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. Journal of Econometrics 132, 195–229.

Corradi, V., Swanson, N.R., 2014. Testing for structural stability of factor augmented forecasting models. Journal of Econometrics 182, 100–118.

Csörgő, M., Horváth, L., 1997. Limit theorems in change-point analysis. volume 18. John Wiley & Sons.

De Haan, L., Ferreira, A., 2007. Extreme value theory: an introduction. Springer Science & Business Media.

Dette, H., Gösmann, J., 2020. A likelihood ratio approach to sequential change point detection for a general class of parameters. Journal of the American Statistical Association 115, 1361–1377.

Eichinger, B., Kirch, C., 2018. A mosum procedure for the estimation of multiple random change points. Bernoulli 24, 526–564.

Fedotenkov, I., 2013. A bootstrap method to test for the existence of finite moments. Journal of Nonparametric Statistics 25, 315–322.

Gasull, A., Jolis, M., Utzet, F., 2015. On the norming constants for normal maxima. Journal of Mathematical Analysis and Applications 422, 376–396.

Gombay, E., Horváth, L., 1996. On the rate of approximations for maximum likelihood tests in change-point models. Journal of Multivariate Analysis 56, 120–152.

Hall, P., 1979. On the rate of convergence of normal extremes. Journal of Applied Probability , 433–439.

Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association 102, 603–617.

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

Han, Y., Chen, R., Zhang, C.H., 2022. Rank determination in tensor factor model. Electronic Journal of Statistics 16, 1726–1803.

He, Y., Kong, X.b., Trapani, L., Yu, L., 2023. One-way or two-way factor model for matrix sequences? Journal of Econometrics *forthcoming* .

Hopke, P.K., Leung, M., Li, N., Navasca, C., 2011. Block tensor decomposition for source apportionment of air pollution. arXiv preprint arXiv:1110.4133 .

Horváth, L., Hušková, M., Kokoszka, P., Steinebach, J., 2004. Monitoring changes in linear models. Journal of Statistical Planning and Inference 126, 225–251.

Horváth, L., Kokoszka, P., Steinebach, J., 2007. On sequential detection of parameter changes in linear regression. Statistics and Probability Letters 80, 1806–1813.

Horváth, L., Miller, C., Rice, G., 2020. A new class of change point test statistics of rényi type. Journal of Business & Economic Statistics 38, 570–579.

Horváth, L., Trapani, L., 2019. Testing for randomness in a random coefficient autoregression. Journal of Econometrics 209, 338–352.

Horvath, L., Trapani, L., 2022. Changepoint detection in random coefficient autoregressive models. Journal of Business and Economic Statistics *forthcoming* .

Komlós, J., Major, P., Tusnády, G., 1975. An approximation of partial sums of independent R.V.'s and the sample DF.I. Z. Wahrscheinlichkeitstheorie und verwandte Gebiete 32, 111–131.

Komlós, J., Major, P., Tusnády, G., 1976. An approximation of partial sums of independent R.V.'s and the sample DF.II. Z. Wahrscheinlichkeitstheorie und verwandte Gebiete 34, 33–58.

Koren, Y., Bell, R., Volinsky, C., 2009. Matrix factorization techniques for recommender systems. Computer 42, 30–37.

Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: inference for the number of factors. Annals of Statistics 40, 694–726.

Lavielle, M., 2005. Using penalized contrasts for the change-point problem. Signal processing 85, 1501–1510.

Li, Q., Cheng, G., Fan, J., Wang, Y., 2018. Embracing the blessing of dimensionality in factor models. Journal of the American Statistical Association 113, 380–389.

Liu, X., Chen, E.Y., 2022. Identification and estimation of threshold matrix-variate factor models. Scandinavian Journal of Statistics .

Massacci, D., 2021. Testing for regime changes in portfolios with a large number of assets: A robust approach to factor heteroskedasticity. Journal of Financial Econometrics *forthcoming* .

Massacci, D., Sarno, L., Trapani, L., 2021. Factor models with downside risk. Available at SSRN .

Müller, U.K., 2013. Risk of bayesian inference in misspecified models, and the sandwich covariance matrix. Econometrica 81, 1805–1849.

Trapani, L., 2016. Testing for (in)finite moments. Journal of Econometrics 191, 57–68.

Trapani, L., 2018. A randomized sequential procedure to determine the number of factors. Journal of the American Statistical Association 113, 1341–1349.

Wang, D., Liu, X., Chen, R., 2019. Factor models for matrix-valued high-dimensional time series. Journal of Econometrics 208, 231–248.

Wang, T., Samworth, R.J., 2018. High dimensional change point estimation via sparse projection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80, 57–83.

Wang, Y., 2023. Identification and Estimation of Parameter Instability in a High Dimensional Approximate Factor Model. Technical Report.

Yamamoto, Y., Tanaka, S., 2015. Testing for factor loading structural change under common breaks. Journal of Econometrics 189, 187–206.

Yu, L., He, Y., Kong, X., Zhang, X., 2022. Projected estimation for large-dimensional matrix factor models. Journal of Econometrics 229, 201–217.

Yong He
Institute for Financial Studies,
Shandong University,
Jinan, 250100, China
E-mail: heyong@sdu.edu.cn

Xinbing Kong
Nanjing Audit University,
Nanjing, 211815, China
E-mail: xinbingkong@126.com

Lorenzo Trapani
University of Leicester Business School,
Leicester, UK
E-mail: lt285@leicester.ac.uk

Long Yu
School of Statistics and Management,
Shanghai University of Finance and Economics,
200433, Shanghai
E-mail: yulong@mail.shufe.edu.cn