Sequential testing for structural stability in approximate factor models

Matteo Barigozzi[1], Lorenzo Trapani[2]

Abstract

We develop a monitoring procedure to detect changes in a large approximate factor model. Letting r be the number of common factors, we base our statistics on the fact that the $(r+1)$-th eigenvalue of the sample covariance matrix is bounded under the null of no change, whereas it becomes spiked under changes. Given that sample eigenvalues cannot be estimated consistently under the null, we randomise the test statistic, obtaining a sequence of *i.i.d* statistics, which are used for the monitoring scheme. Numerical evidence shows a very small probability of false detections, and tight detection times of change-points.

Keywords: large factor model, change-point, sequential testing, randomised tests.

[1]Department of Statistics, London School of Economics and Political Science.
Email: m.barigozzi@lse.ac.uk (corresponding author).
[2]School of Economics, The University of Nottingham
Email: Lorenzo.Trapani1@nottingham.ac.uk

We wish to thank the participants to the International Conference in Memory of Carlo Giannini (Bergamo, November 25-26, 2016); to the CMStatistics 2016 conference, (Seville, December 9-11, 2016); to the ECARES Seminar Series (Bruxelles, February 23, 2017); to the Cass Business School Faculty of Finance Workshop (London, March 6, 2017); to the LSE Econometrics and Statistics seminar (London, March 10, 2017); to the Department of Economics Seminar Series at the University of Connecticut (March 31, 2017); to the New York Camp Econometrics XII conference (Syracuse University, April 7-9, 2017); and to the European Meeting of Statisticians (University of Helsinki, July 24-28, 2017).

1. Introduction

In this paper, we investigate the issue of testing for the stability of a large factor model:

$$X_{i,t} = a_i' f_t + u_{i,t}, \tag{1}$$

where $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$ is a panel of N time series observed for T periods; a_i and f_t are latent vectors of loadings and factors, respectively, both of dimension $r < N$ and representing the "signal" component of the data, as opposed to the idiosyncratic "noise" $u_{i,t}$. In particular, we focus on the *sequential monitoring* of the stability of (1) - that is, we propose a test to check whether there are any breaks in (1) as new data come in. Factor models have been paid significant attention in virtually all applied sciences, as a tool to reduce dimensionality while preserving the information content of a large dataset. In particular, in the context of social sciences and economics, the use of factor models has been popularised by the seminal paper by Chamberlain and Rothschild [19]; thereafter, factor models have acquired a huge popularity in various applications, such as business cycle analysis, asset pricing and economic monitoring and forecasting – see the review by Stock and Watson [61] for a comprehensive list of references.

Model (1) is usually characterized by the identifying assumption that, as $N \to \infty$, the covariance matrix of $\{X_{i,t}\}_{i=1}^N$ has r spiked eigenvalues diverging to infinity, while the remaining ones stay bounded for any N. Numerous contributions have developed a full-fledged inferential theory for (1) under general assumptions, such as weak serial and cross-correlation of the error terms $u_{i,t}$. In particular, in the case of stationary data the estimation by means of principal component analysis of the "signal" part of (1) has been developed for high-dimensional, i.e. large N, data, e.g. by Bai [6] and Fan *et al.* [29]. The literature has also produced many results on the determination of the number of common factors r – see, *inter alia*, Bai and Ng [8], Alessi *et al.* [2], Onatski [52], Ahn and Horenstein [1], and Trapani [62]. The factors in equation (1) have also proven to be very effective to forecast large datasets, overcoming the curse of dimensionality issue – see e.g. Stock and Watson [58]. Extensions to the case in which the common factors f_t are explicitly allowed to have a linear process representation, have been studied also – see e.g. Forni *et al.* [31].

In comparison with this huge body of literature, the issue of testing for the structural stability of (1) can be still considered underdeveloped, with some notable exceptions. Indeed, Stock and Watson [58] and Bates *et al.* [13] argue that, at least in the presence of "small" breaks and a constant number of factors, inference on the factor space is not hampered, thus making the change-point problem less compelling than in other contexts. Nevertheless, stylised facts show that in many applications the assumptions of a negligible break size and a stable number of factors are not, in general, correct. Most importantly, it has been argued that, in presence of a crisis, co-movements become stronger, which may suggest that the economy is driven by a different number of factors than in quieter periods – see e.g. Stock and Watson [60], Cheng *et al.* [22] and Li *et al.* [47]. In such cases, the impact of a change-point is bound to invalidate standard inference and subsequent applications such as forecasting. Recently, the literature has proposed a series of tests for the in-sample detection of breaks in factor structures: examples include the works by Breitung and Eickmeier [16], Chen *et al.* [20], Han and Inoue [33], Corradi and Swanson [25], Yamamoto and Tanaka [65], Cheng *et al.* [22], Baltagi *et al.* [10], Massacci [48] and Barigozzi *et al.* [12].

Sequential detection of breaks in (1) is important for at least four reasons. First, the general motivation put forward by Chu *et al.* [23] holds true in the context of factor models also: it is important to verify whether a model, which has been valid thus far, is still capable of adequately approximate the behaviour of new data. Second, the aforementioned (substantial) empirical evidence that factor structures do tend to change over time, especially in presence of a crisis, illustrates the importance of a timely detection of such changes. Third, inference on factor models can be severely marred by the presence of a break (see the comments in Baltagi *et al.* [10]), which again shows the importance of detecting a break in real time, rather than realising this *a posteriori* after inference has been carried out and employed, e.g. for the purpose of forecasting. Finally, in the context of economics and finance, data are collected and made available automatically, so that the cost of monitoring is almost negligible, especially if compared with the potential costs of employing a model which is no longer valid. Sequential detection of breaks in a univariate or small dimensional, i.e. finite N, setting has been studied e.g. in Lai [45], Chu *et al.* [23], Aue and Horváth [4], Horváth *et al.* [37], Andreou and Ghysels [3], Horváth *et al.* [38], Brodsky [17], Aue *et al.* [5], Kirch and Tadjuidje Kamgaing [41], and Groen *et al.* [32].

1.1. Hypotheses of interest and main results of the paper

There are several possible ways in which model (1) may undergo a change at a point in time τ; however, despite such a wide variety, in all cases it may be argued that a change in the factor structure of the data will result in a change in the covariance matrix of $\{X_{i,t}\}_{i=1}^N$. More specifically, since common factors determine the presence and number of spiked eigenvalue of the covariance of $\{X_{i,t}\}_{i=1}^N$ (defined as eigenvalues which are not bounded, but grow with the dimension of the dataset), it is natural to investigate whether a change has occurred in the factor structure of (1) by verifying whether changes have occurred in the spectrum of the covariance matrix. Formally, in this paper we test for the null hypothesis that the factor structure does not change, viz.:

$$H_0 : X_{i,t} = \sum_{j=1}^{r} a_{ij} f_{j,t} + u_{i,t}, \quad 1 \leq t \leq T.$$

As far as alternatives are concerned, we focus on two different possible breaks at a point in time τ: (1) changes in the loadings attached to one or more common factor:

$$H_{A,1} : \begin{cases} X_{i,t} = \sum_{j=1}^{r} a_{ij} f_{j,t} + u_{i,t} & \text{for} \quad 1 \leq t < \tau \\ X_{i,t} = \sum_{j=1}^{r} \widetilde{a}_{ij} f_{j,t} + u_{i,t} & \quad\quad\quad \tau \leq t \leq T \end{cases}, \tag{2}$$

where $\widetilde{a}_{ij} \neq a_{ij}$ for all i and at least one value of j, and (2) the appearance of $q \geq 1$ new factors:

$$H_{A,2} : \begin{cases} X_{i,t} = \sum_{j=1}^{r} a_{ij} f_{j,t} + u_{i,t} & \text{for} \quad 1 \leq t < \tau \\ X_{i,t} = \sum_{j=1}^{r} a_{ij} f_{j,t} + \sum_{j=1}^{q} b_{ij} g_{j,t} + u_{i,t} & \quad\quad\quad \tau \leq t \leq T \end{cases}. \tag{3}$$

Hypothesis $H_{A,1}$ is the typical case considered in all the above cited literature on change-points in factor models. A consequence of (2) is that, under the alternative, a model with r common factors and changing loadings can be re-written as a model with a total number of factors ranging between $r+1$ and $2r$ common factors, defined as the original common factors multiplied by a pre- and post-break dummy variable. This key property is heavily

exploited in the literature. On the other hand hypothesis $H_{A,2}$ has received less attention from the literature – see for example Cheng *et al.* [22] and Barigozzi *et al.* [12]. Whilst in this paper we mainly focus on $H_{A,1}$ and $H_{A,2}$, other alternatives, as disappearing factors or less pervasive changes in the loadings, can also be accommodated in our framework – see the discussion in Section 4.

We show that, under both $H_{A,1}$ and $H_{A,2}$, the $(r+1)$-th largest eigenvalue of the covariance matrix of $\{X_{i,t}\}_{i=1}^N$ becomes unbounded at time τ, passing to infinity as fast as the sample size N. Conversely, it stays bounded under the null of no break. Thus, we base our test on the estimated $(r+1)$-th eigenvalue of the sample covariance matrix of $\{X_{i,t}\}_{i=1}^N$ computed using a rolling window. Although using the sample eigenvalues of the sample covariance matrix for testing is not uncommon in the context of factor models (Onatski [52], Trapani [62]), in our context such an approach is fraught with difficulties. The main issue is that, under the null of no break, the $(r+1)$-th sample eigenvalue does not have a known distribution, and indeed it cannot even be estimated consistently: as Wang and Fan [63] explain, there is too much noise (due to N being large) to be able to identify the small signal coming from a bounded eigenvalue.

Given that the only thing we know is that the $(r+1)$-th sample eigenvalue may be bounded or unbounded, we propose to use a randomised test in order to regularize the problem. Randomisation is a widely employed approach, dating back at least to Pearson [56]; various authors have employed different ways of introducing randomness into a statistic – see e.g. Corradi and Swanson [24], Bandi and Corradi [11], and Trapani [62]. Our methodology is based on the same approach, but with a different scope. In essence, the approach which we propose takes, at each point in time t, the $(r+1)$-th sample eigenvalue as input, and returns, as output, an *i.i.d.* sequence, with known (asymptotic) distribution, first and second moments that can be approximated with a negligible error, and finite moments up to any order. Such sequence is then used to replace the $(r+1)$-th sample eigenvalue in the construction of the monitoring process, thus allowing us to use the standard asymptotic theory already developed for partial sum processes of *i.i.d.* sequences – see Horváth *et al.* [37] and Kirch and Tadjuidje Kamgaing [41]. Although our results are derived conditionally on the sample (see the comments in Section 3 on the meaning of randomisation under sample conditioning), we construct a monitoring procedure which falsely identifies a break under the null with probability smaller than a prescribed level, and which identifies a break with probability one when this is present. This is a desirable feature of sequential testing since as more data come in the probability of type I errors is anyway likely to increase – see for example the comments in Chapter 9 by Sen [57]. Indeed, numerical evidence suggests that our procedure works extremely well, with a short delay in finding breaks. In principle, our test can be applied also under more general circumstances, including the presence of weak factors or less pervasive loadings changes, the case of heteroskedastic idiosyncratic components, and the disappearance of one or more factors. All these extensions are discussed in Sections 4 and 7.

The rest of the paper is organised as follows. In Section 2 we spell out the main assumptions, and we study the inference on the $(r+1)$-th eigenvalue of the covariance matrix. Section 3 discusses the construction of the test statistic, including the double randomisation procedure and all the relevant intermediate results. Some straightforward extensions of our framework to more general circumstances are discussed in Section 4. Numerical evidence from Monte Carlo experiments and a real data application on US

industrial production monthly data are given in Sections 5 and 6, respectively. Section 7 discusses further possible extensions and concludes. All proofs are in the Appendix.

NOTATION. We let $C_0, C_1, ...$ denote generic, finite positive constants that do not depend on the sample size, and whose value may change from line to line; "$\rightarrow$" denotes the ordinary limit; orders of magnitude for an a.s. convergent sequence (say s_T) are denoted as $O_{a.s.}(T^{\varsigma})$ and $o_{a.s.}(T^{\varsigma})$ when, for some $\epsilon > 0$ and $\tilde{T} < \infty$, $P\left[|T^{-\varsigma} s_T| < \epsilon \text{ for all } T \geq \tilde{T}\right] = 1$ and $T^{-\varsigma} s_T \rightarrow 0$ a.s., respectively; $I_A(x)$ is the indicator function of a set A. Finally, we assume without loss of generality that all random variables and processes are defined on a common probability space $(\Omega, \mathcal{F}, P)$ with outcomes $\omega \in \Omega$.

2. Assumptions and preliminary theory

Consider the factor model in (1), where we now make explicit the possibility of changes over time in the "signal" component

$$X_{i,t} = a_i'(t) f_t + u_{i,t}, \quad 1 \leq i \leq N,\ 1 \leq t \leq T. \tag{4}$$

We use $r(t)$ to denote the number of factors at a given time t, i.e. the vectors of loadings $a_i(t)$ and of factors f_t have dimension $r(t)$. Consider also the matrix form of (4):

$$X_t = A(t) f_t + u_t, \quad 1 \leq t \leq T, \tag{5}$$

where, $A(t) = [a_1(t)|...|a_N(t)]'$ is the $N \times r(t)$ loadings matrix and $u_t = [u_{1,t}, ..., u_{N,t}]'$ is the idiosyncratic component. Under $H_{A,1}$ and $H_{A,2}$ (see also (2) and (3)), we define $\widetilde{a}_i = (\widetilde{a}_{i,1}, ..., \widetilde{a}_{i,r})'$ and $\widetilde{A} = [\widetilde{a}_1|...|\widetilde{a}_N]'$, and $b_i = (b_{i,1}, ..., b_{i,q})'$ and $B = [b_1|...|b_N]'$. Using this notation, we are interested in testing the null-hypothesis

$$H_0 : A(t) = A, \quad 1 \leq t \leq T,$$

versus the alternatives

$$H_{A,1} : \begin{cases} A(t) = A & \text{for} \quad 1 \leq t < \tau \\ A(t) = \widetilde{A} & \quad\ \ \tau \leq t \leq T \end{cases},$$

and

$$H_{A,2} : \begin{cases} A(t) = A & \text{for} \quad 1 \leq t < \tau \\ A(t) = [A|B] & \quad\ \ \tau \leq t \leq T \end{cases}$$

We define the covariance matrix of the data at time t as $\Sigma_X(t) = E(X_t X_t')$, assuming for simplicity, and without loss of generality, that X_t has zero-mean. Consider the (population) rolling covariance matrix

$$\Sigma_m(t) = \frac{1}{m} \sum_{k=t-m+1}^{t} \Sigma_X(k), \quad m \leq t \leq T, \tag{6}$$

and its sample counterpart

$$\widehat{\Sigma}_m(t) = \frac{1}{m} \sum_{k=t-m+1}^{t} X_k X_k', \quad m \leq t \leq T. \tag{7}$$

Based on (6) and (7), in what follows m will denote our sample size when estimating the model; hence, our asymptotics is for $m \to \infty$. We assume that for the first m periods no change-point is present and we have $r(t) = r$ factors for all $t \leq m$. Moreover, for simplicity, we also assume that our monitoring procedure will last until $T > m$. Therefore, the total number of observations T includes both the estimation and the monitoring period. Note that, in real applications, the monitoring may be expected to go on indefinitely, so that $T \to \infty$.

We start with the following assumption.

Assumption 1. *It holds that (i)* $E(X_{i,t}) = 0$ *for all* $1 \leq i \leq N$ *and* $1 \leq t \leq T$*; (ii)* $E(f_{j,t}u_{i,t}) = 0$ *for all* i, j, t*; (iii)* $r(t) = r$ *for* $1 \leq t \leq m$*; (iv)* $r(t) < N$ *and finite for* $1 \leq t \leq T$ *and for all* $N \in \mathbb{N}$*.*

Parts *(i)* and *(ii)* of the assumption are made only for convenience and could be relaxed. Clearly from part *(iii)* we have that, in presence of breaks, the change-point location τ is such that $\tau > m$. Finally, part *(iv)* is a reasonable requirement for the number of factors to be finite at any point in time. Note that, under H_0 and $H_{A,1}$ we have that $r(t) = r$ for all $m \leq t \leq T$, while under $H_{A,2}$ $r(t) = r$ for $1 \leq t < \tau$ and $r(t) = (r + q)$ for $\tau \leq t \leq T$.

By Assumption 1 the covariance is decomposed as

$$\Sigma_X(t) = A(t)\,\Sigma_F(t)\,A(t)' + \Sigma_u(t),$$

having defined $\Sigma_F(t) = E(f_t f_t')$ and $\Sigma_u(t) = E(u_t u_t')$. Henceforth, we denote the k-th largest eigenvalue of $\Sigma_m(t)$ as $\lambda^{(k)}(t)$, the k-th eigenvalue of $A(t)\,\Sigma_F(t)\,A(t)'$ as $\gamma^{(k)}(t)$; and, finally, the k-th eigenvalue of $\Sigma_u(t)$ as $\omega^{(k)}(t)$; similarly, we denote the k-th largest eigenvalue of $\widehat{\Sigma}_m(t)$ as $\widehat{\lambda}^{(k)}(t)$.

In order to derive our results on the population and sample eigenvalues, we make the following assumptions.

Assumption 2. *It holds that (i)* $\underline{C}_k(t)N \leq \gamma^{(k)}(t) \leq \overline{C}_k(t)\,N$ *for all* $1 \leq k \leq r(t)$*, and* $0 < \underline{C}_k(t) \leq \overline{C}_k(t) < \infty$ *and for* $m \leq t \leq T$*; (ii)* $\omega^{(k)}(t) \leq C_0$ *for all* $1 \leq k \leq N$ *and* $m \leq t \leq T$*.*

Assumption 3. *It holds that (i)* $E\left|X_{i,t}\right|^{4+\epsilon} \leq C_0$ *for all* $1 \leq i \leq N$*,* $1 \leq t \leq T$ *and some* $\epsilon > 0$*; (ii)* $E\left[\max_{t_0 \leq \tilde{t} \leq t_0+m-1}\left|\sum_{t=t_0}^{\tilde{t}} X_{h,t}X_{j,t} - E\left(X_{h,t}X_{j,t}\right)\right|^2\right] \leq C_1 m$ *for all* $1 \leq h, j \leq N$ *and* $1 \leq t_0 \leq T - m + 1$*.*

Assumption 2 is typical of high-dimensional factor analysis and is analogous to the assumptions in Chamberlain and Rothschild [19] and Forni *et al.* [31]. In particular, as far as the non-zero $\gamma^{(k)}(t)$'s are concerned, part *(i)* of the assumption requires that they diverge to positive infinity, as $N \to \infty$, at a rate N.

Equivalently, we could follow Bai and Ng [8] and Fan *et al.* [29] and require the more primitive assumptions that $\Sigma_F(t)$ is positive definite (which entails that common factors are identified), and that $N^{-1}A(t)'A(t)$ tends to a positive definite matrix. This is tantamount to assuming that $\gamma^{(k)}(t)$ passes to infinity at a rate N. Indeed, consider - for the sake of the notation - the case of constant loadings, viz. $A(t) = A$, and constant covariance matrix for the common factors, viz. $\Sigma_F(t) = \Sigma_F$; then, using Theorem 7 in Merikoski and Kumar

[49]

$$N\nu^{(\min)}\left(\Sigma_F\right)\nu^{(k)}\left(\frac{A'A}{N}\right) \leq \nu^{(k)}\left(A\Sigma_F A'\right) \leq N\nu^{(\max)}\left(\Sigma_F\right)\nu^{(k)}\left(\frac{A'A}{N}\right),$$

where $\nu^{(k)}(\cdot)$ denotes the k-th largest eigenvalue of a matrix. When following the same reasoning in the presence of change-points, the above result provides a link between $\gamma^{(k)}(t)$ and the k-th largest eigenvalue of $N^{-1}A'(t)A(t)$. Note that it is also possible to assume that $\gamma^{(k)}(t) \to \infty$ as $N \to \infty$ at a slower rate than N, which is known as having "weak factors"; we discuss this case in Section 4.1.

As far as the $\omega^{(k)}(t)$'s are concerned, in part *(ii)* of the assumption, the same condition could be derived from the assumptions in Fan *et al.* [29] – see also Bai and Ng [8]. Note also that we do not require the $\omega^{(k)}(t)$'s to be constant over t: unconditional heteroskedasticity is allowed for, in principle – see also the comments in Section 4. Assumption 2 determines the behaviour of the population eigenvalues of $\Sigma_m(t)$. In particular, at $t = m$, by Weyl's inequality we have that $\lambda^{(k)}(m) \geq C_k(m)N$ for $1 \leq k \leq r$, while $\lambda^{(k)}(m) \leq C_0$ for $r+1 \leq k \leq N$. This condition implies the existence of an eigen-gap which allows us to identify r in the pre-break sample.

As far as Assumption 3 is concerned, part *(ii)* is a high-level condition which, in essence, poses a constraint on the amount of serial correlation that one can have in the process $\{X_{h,t}X_{j,t}\}_{t=1}^T$ and therefore, albeit indirectly, in $\{X_{i,t}\}_{t=1}^T$. In general, this assumption is satisfied by any linear process with summable fourth cumulants (see e.g., Hannan [34], Theorem 6, page 210). Some examples under which Assumption 3 holds are reported in Trapani [62] and include the case of stationary, causal processes – see Wu [64]. This family of processes in turn includes several popular examples such as Volterra series and ARCH/GARCH processes, thus allowing for the case of conditional heteroskedasticity.

Finally, note that Assumptions 2 and 3 allow for some degree of cross-sectional and serial dependence in the panel of idiosyncratic components, $\{u_{i,t},\, 1 \leq i \leq N,\, 1 \leq t \leq T\}$; thus, (1) defines an "approximate" factor model, as opposed to an "exact" one, which would require cross-sectionally and serially *i.i.d.* errors.

The following result characterizes the behaviour of the $(r+1)$-th eigenvalue of $\Sigma_m(t)$.

Lemma 1. *Under Assumptions 1 and 2, it holds that*

$$\lambda^{(r+1)}(t) \leq C_0, \quad m \leq t \leq T, \ \textit{under } H_0. \tag{8}$$

Further, it holds that

$$\lambda^{(r+1)}(t) \begin{cases} \leq C_0 & m \leq t < \tau, \\ \geq C_1 \min\{\frac{t-\tau+1}{m}, \frac{\tau+m-t-1}{m}\}N & \tau \leq t < \tau+m-1, \\ \leq C_0 & \tau+m-1 \leq t \leq T, \quad \textit{under } H_{A,1}, \end{cases} \tag{9}$$

$$\lambda^{(r+1)}(t) \begin{cases} \leq C_0 & m \leq t < \tau, \\ \geq C_1 \frac{t-\tau+1}{m} N & \tau \leq t < \tau+m-1, \\ \geq C_1 N & \tau+m-1 \leq t \leq T, \quad \textit{under } H_{A,2}. \end{cases} \tag{10}$$

The sample counterpart to Lemma 1 is the following result, derived in Trapani [62].

Lemma 2. *Under Assumptions 1 and 3, it holds that*

$$\widehat{\lambda}^{(r+1)}(t) = \lambda^{(r+1)}(t) + O_{a.s.}\left(\frac{N}{m^{1/2}} l(m,N)\right), \quad m \leq t \leq T, \tag{11}$$

where

$$l(m,N) = (\ln N)^{1+\epsilon} (\ln m)^{\frac{1+\epsilon}{2}},$$

for any $\epsilon > 0$.

Lemma 2 provides a strong rate for the estimation error $(\widehat{\lambda}^{(r+1)}(t) - \lambda^{(r+1)}(t))$, which is valid for any combination of N and m, and indeed for all estimated eigenvalues, $\widehat{\lambda}^{(k)}(t)$ for $1 \leq k \leq N$. The lemma does not require any assumption on $\lambda^{(k)}(t)$: some of these may be non-distinct, non well-separated, or even equal to zero. Equation (11) states that the estimation error can be quite large. It is, however, comparatively small for the spiked eigenvalues, which, by Assumption 2, are of order N. Conversely, the error term in (11) can be quite large for the bounded eigenvalues; in this case, the rate is probably not the sharpest one, although it suffices for the construction of the monitoring procedure. The result of Lemma 2 can also be compared with the results from Random Matrix literature for spiked covariance models where however $\lambda^{(k)}$ is finite for all $1 \leq k \leq N$ and $N \in \mathbb{N}$ – see e.g. El Karoui [28], Paul [55], Johnstone and Lu [39], Jung and Marron [40], Benaych-Georges and Nadakuditi [14, 15], Bai and Yao [9], and Onatski *et al.* [54].

3. Testing procedure and asymptotics

In this section, we propose an algorithm to "regularise" the behaviour of the eigenvalues so as to be able to construct a monitoring procedure. As a consequence of Lemmas 1 and 2, we are unable to use $\widehat{\lambda}^{(r+1)}(t)$, due to the lack of a known limiting distribution under the null, and of the dependence structure across t. We therefore propose a randomisation algorithm, whose output is a sequence of *i.i.d.* random variables with finite moments of arbitrarily high order and, under the null, (asymptotically) chi-square distributed. We subsequently employ (the standardised version of) such random variables to construct a partial sum process, which we use as the relevant test statistic in an analogous way as Horváth *et al.* [37] and Horváth *et al.* [38].

3.1. The randomisation algorithm

Define $\delta \in (0,1)$ such that

$$\delta \begin{cases} > 0 \\ > 1 - \frac{1}{2}\frac{\ln m}{\ln N} \end{cases} \text{ according as } \begin{array}{l} N \leq m^{1/2} \\ N > m^{1/2} \end{array}; \tag{12}$$

note that the choice of δ is uniquely determined by N and m, with no need to estimate it. We consider the statistic

$$\phi_{N,m}(t) = g\left(\frac{N^{-\delta}\widehat{\lambda}^{(r+1)}(t)}{\frac{1}{N}\sum_{k=1}^{N}\widehat{\lambda}^{(k)}(t)}\right), \quad m \leq t \leq T, \tag{13}$$

where $g(\cdot)$ is a monotonically increasing function such that $g(0) = 0$ and $\lim_{x\to\infty} g(x) = \infty$; in this paper, we use $g(a) = a$, but other choices are also possible. The denominator in (13) makes the argument of $g(\cdot)$ scale invariant.

The quantity δ, defined in (12), plays a very important role in the remainder of the paper. Based on Lemma 2, it can be expected that $\widehat{\lambda}^{(r+1)}(t)$ may diverge to positive infinity even when $\lambda^{(r+1)}(t)$ is bounded; in this case, the divergence rate is $O\left(Nm^{-1/2}\right)$, modulo the logarithmic terms. On the other hand, $\widehat{\lambda}^{(r+1)}(t)$ diverges at the faster rate $O(N)$ under the alternative. The purpose of δ is to annihilate the estimation error: based on (12), it can be seen that N^{δ} is larger than $Nm^{-1/2}l(m,N)$: thus, under the null of no break, it can be expected that $N^{-\delta}\widehat{\lambda}^{(r+1)}(t)$ will drift to zero. Under the alternative, it still passes to infinity (since $\delta<1$), albeit at a slower rate than $\lambda^{(r+1)}(t)$ itself. Note that this would hold also for very large values of N: indeed, no restriction is required between the relative rate of divergence of N and m as they pass to infinity, and one could also allow for $N=\exp(m)$; in this case, after some algebra it can be shown that $\delta\in\left(1-\frac{1}{2}\frac{\ln m}{m},1\right)$, which still yields that $N^{-\delta}\widehat{\lambda}^{(r+1)}(t)$ drifts to zero or diverges to infinity according as the null or the alternative is true.

On account of the comments above, and of Lemmas 1 and 2, it holds that

$$\lim_{N,m\to\infty}\phi_{N,m}(t)=g(0)=0,\quad \text{w.p. 1, when } N^{-\delta}\lambda^{(r+1)}(t)\to 0,$$
$$\lim_{N,m\to\infty}\phi_{N,m}(t)=g(\infty)=\infty,\quad \text{w.p. 1, when } N^{-\delta}\lambda^{(r+1)}(t)\to \infty.$$

We therefore have that

$$\lim_{N,m\to\infty}\phi_{N,m}(t)=0,\quad m\le t\le T,\quad \text{under } H_0.$$

Henceforth, we define $t^*_{N,m}$ as the point in time such that $t^*_{N,m}\ge\tau$ and

$$\lim_{N,m\to\infty}\frac{N^{1-\delta}}{m}\left(t^*_{N,m}-\tau+1\right)=\infty. \tag{14}$$

Similarly, we define the point in time $t^{**}_{N,m}\le\tau+m-1$ such that

$$\lim_{N,m\to\infty}\phi_{N,m}(t)=\begin{cases}0 & m\le t<\tau,\\ \infty & t^*_{N,m}\le t<t^{**}_{N,m},\\ 0 & \tau+m-1\le t\le T,\end{cases},\quad \text{under } H_{A,1}.$$

Clearly

$$\lim_{N,m\to\infty}\phi_{N,m}(t)=\begin{cases}0 & m\le t<\tau,\\ \infty & t^*_{N,m}\le t<\tau+m-1,\end{cases},\quad \text{under } H_{A,2};$$

Under $H_{A,1}$ for $\tau\le t<t^*_{N,m}$, $\phi_{N,m}(t)$ is growing from 0 to ∞, and viceversa for $t^{**}_{N,m}\le t<\tau+m-1$, while under $H_{A,2}$ for $\tau\le t<t^*_{N,m}$, $\phi_{N,m}(t)$ is growing from 0 to ∞. Therefore, $t^*_{N,m}$ represents the first point in time in which we can hope to detect the change, hence is a lower bound for the delay in detection, while under $H_{A,1}$ $(t^{**}_{N,m}-1)$ represents the last point in time in which we can hope to detect the change. In light of (14) and the results that follow, we show in Section 3.3 below that $t^*_{N,m}$ is at least of order $m^{1/2}$ regardless of the values of m and N.

Given that the results above entail that we only have rates for $\phi_{N,m}(t)$, we propose a to use a randomised version of it, built according to the following steps.

Step A1. At each given $t \geq m$, generate an *i.i.d.* sample $\{\xi_j(t)\}_{j=1}^{R}$ with common distribution G_ϕ such that $G_\phi(0) \neq 0$ or 1.

Step A2. For any u drawn from a distribution $F_\phi(u)$, define

$$\zeta_j(u;t) = I\left[\xi_j(t) \leq u\phi_{N,m}^{-1}(t)\right].$$

Step A3. Compute

$$\vartheta(u;t) = \frac{1}{\sqrt{R}} \sum_{j=1}^{R} \frac{\zeta_j(u;t) - G_\phi(0)}{\sqrt{G_\phi(0)[1 - G_\phi(0)]}}.$$

Step A4. Compute

$$\Theta_t = \int_{-\infty}^{+\infty} |\vartheta(u;t)|^2 \, dF_\phi(u).$$

Although the details of the behaviour of Θ_t under the null and the alternative are spelt out later on, a heuristic preview of the main argument may be helpful. In essence, under the alternative the Bernoulli random variable $\zeta_j(u;t)$ should be equal to 1 or 0 with probability $G_\phi(0)$ and $1 - G_\phi(0)$ respectively, and thus have mean $G_\phi(0)$. In this case, when constructing $\vartheta(u;t)$, a Central Limit Theorem holds and therefore we expect Θ_t to have a chi-square distribution. On the other hand, under the null $\zeta_j(u;t)$ should be (heuristically) 0 or 1 with probability 0 or 1 (depending on the sign of u) - thus, its mean should be different than $G_\phi(0)$ (and equal to 0 or 1 depending on the sign of u) and a Law of Large Numbers should hold. Note that, by construction, conditionally on the sample the sequence $\{\Theta_t\}_{t=m}^{T}$ is independent across t. In order to study Θ_t, we need the following assumptions.

Assumption 4. *It holds that: (i) $G_\phi(\cdot)$ has a bounded density; (ii) $\int_{-\infty}^{+\infty} u^2 dF_\phi(u) < \infty$; (iii) $F_\phi(0) < 1$.*

Assumption 5. *It holds that, as $\min(N, m, R) \to \infty$:*

$$\text{(i)} \quad R^{1/2}\left[g\left(N^{1-\delta}\frac{t-\tau+1}{m}\right)\right]^{-1} \to 0, \quad \begin{array}{ll} \text{under } H_{A,1}, & \text{for } t_{N,m}^{*} \leq t < t_{N,m}^{**}, \\ \text{under } H_{A,2}, & \text{for } t_{N,m}^{*} \leq t \leq T; \end{array}$$

(ii) $R^{1/2}\left[g\left(N^{1-\delta}\right)\right]^{-1} \to 0$ under $H_{A,1}$, for $\tau + m - 1 \leq t \leq T$.

Considering Assumption 4, G_ϕ can be chosen as the standard normal distribution, and F_ϕ as a discrete uniform distribution. Assumption 5 provides a selection rule for R.

Let now P^* represent the conditional probability with respect to $\{X_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$; "$\overset{\mathcal{D}^*}{\to}$" and "$\overset{P^*}{\to}$" denote, respectively, conditional convergence in distribution and in probability according to P^*.

Theorem 1. *Under Assumptions 1-5, as $\min(N, m, R) \to \infty$, it holds that*

$$\Theta_t \overset{\mathcal{D}^*}{\to} \chi_1^2, \quad \begin{array}{ll} \text{under } H_{A,1}, & \text{for } t_{N,m}^{*} \leq t < t_{N,m}^{**}, \\ \text{under } H_{A,2}, & \text{for } t_{N,m}^{*} \leq t \leq T, \end{array} \tag{15}$$

for almost all realisations of $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$.
Under Assumptions 1-4, as $\min(N, m, R) \to \infty$*, it holds that*

$$\frac{1}{R}\Theta_t \overset{P^*}{\to} \frac{\int_{-\infty}^{+\infty} \left|I_{[0,\infty)}(u) - G_\phi(0)\right|^2 dF_\phi(u)}{G_\phi(0)\left[1 - G_\phi(0)\right]}, \quad \begin{array}{ll} \textit{under } H_0, & \textit{for } \ m \leq t \leq T, \\ \textit{under } H_{A,1}, & \textit{for } \ m \leq t < \tau, \\ & \textit{and } \tau + m - 1 \leq t \leq T, \\ \textit{under } H_{A,2}, & \textit{for } \ m \leq t < \tau, \end{array} \tag{16}$$

for almost all realisations of $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$.

Theorem 1 is an intermediate result: in order to be able to construct a test for the "classical" null of no changes in the factor structure, it is necessary to have a statistic which diverges under the null and is bounded under the alternative. In particular, the behaviour under the null is - clearly - very important to ensure size control of the monitoring procedure. As can be noted, the reason why Θ_t is bounded under the null is because we have constructed a statistic based on randomising the estimated eigenvalue $\widehat{\lambda}^{(r+1)}(t)$. Thus, it can be envisaged that randomising its reciprocal would yield the desired behaviour. Whilst this is theoretically possible, we recommend against it: as Lemma 2 shows, in this case, under the null, the behaviour of Θ_t would be driven by a term proportional to (the inverse of) $N^{-\delta}\frac{N}{m^{1/2}}$: but since this estimate is only an upper bound, and thus not sharp (contrary to the case of randomising $\widehat{\lambda}^{(r+1)}(t)$ directly), it is unclear what the rate of divergence would be in this case.
We therefore propose to randomise Θ_t, with a second randomisation based on

$$\psi_{N,m,R}(t) = h\left(\frac{\Theta_t}{\widetilde{l}(N, m, R)}\right), \quad m \leq t \leq T, \tag{17}$$

where

$$\widetilde{l}(N, m, R) = (\ln N)^{2+\epsilon}(\ln m)^{2+\epsilon}(\ln R)^{2+\epsilon},$$

for some $\epsilon > 0$ - in practice, any small value of ϵ works well.

In (17), the function $h(\cdot)$, similarly to $g(\cdot)$ in (13), is a monotonically increasing function such that $h(0) = 0$ and $\lim_{x\to\infty} h(x) = \infty$; again, we use $h(a) = a$.

Similarly to the case of $\phi_{N,m}(t)$, Theorem 1 entails that

$$\lim_{N,m,R\to\infty} \psi_{N,m,R}(t) = \infty, \quad m \leq t \leq T, \quad \text{under } H_0,$$

and

$$\lim_{N,m,R\to\infty} \psi_{N,m,R}(t) = \begin{cases} \infty & m \leq t < \tau, \\ 0 & t^*_{N,m} \leq t < t^{**}_{N,m}, \\ \infty & \tau + m - 1 \leq t \leq T, \end{cases}, \quad \text{under } H_{A,1},$$

while

$$\lim_{N,m,R\to\infty} \psi_{N,m,R}(t) = \begin{cases} \infty & m \leq t < \tau, \\ 0 & t^*_{N,m} \leq t < \tau + m - 1, \end{cases}, \quad \text{under } H_{A,2}.$$

Consider now the second randomisation.

Step B1. At each given $t \geq m$, generate an *i.i.d.* sample $\left\{\widetilde{\xi}_j(t)\right\}_{j=1}^{W}$ with common distribution G_ψ such that $G_\psi(0) \neq 0$ or 1.

Step B2. For any u drawn from a distribution $F_\psi(u)$, define

$$\widetilde{\zeta}_j(u;t) = I\left[\widetilde{\xi}_j(t) \leq u\psi_{N,m,R}^{-1}(t)\right].$$

Step B3. Compute

$$\gamma(u;t) = \frac{1}{\sqrt{W}} \sum_{j=1}^{W} \frac{\widetilde{\zeta}_j(u;t) - G_\psi(0)}{\sqrt{G_\psi(0)[1 - G_\psi(0)]}}.$$

Step B4. Compute

$$\Gamma_t = \int_{-\infty}^{+\infty} |\gamma(u;t)|^2 \, dF_\psi(u).$$

The following assumptions are needed in order to study the asymptotic behavior of Γ_t; note their similarity with Assumptions 4 and 5.

Assumption 6. *It holds that: (i) $G_\psi(\cdot)$ has a bounded density; (ii) $\int_{-\infty}^{+\infty} u^4 dF_\psi(u) < \infty$; (iii) $F_\psi(0) < 1$.*

Assumption 7. *It holds that, as* $\min(N, m, R, W) \to \infty$

$$W^{1/2}\left[h\left(\frac{R}{\widetilde{l}(N,m,R)}\right)\right]^{-1} \to 0.$$

As above, in Assumption 6 we can choose G_ψ to be the standard normal distribution, and F_ψ to be a discrete uniform distribution. The restrictions in Assumption 7 provide a selection rule for W.

Let $P^\dagger$ represent the conditional probability with respect to $\{X_{i,t}, 1 \leq i \leq N,\ 1 \leq t \leq T\}$ and $\{\xi_j(t), 1 \leq j \leq R,\ m \leq t \leq T\}$; we use the notation "$\overset{\mathcal{D}^\dagger}{\to}$" and "$\overset{P^\dagger}{\to}$" to define, respectively, conditional convergence in distribution and in probability according to $P^\dagger$.

Theorem 2. *Under Assumptions 1-7, as* $\min(N, m, R, W) \to \infty$*, it holds that*

$$\Gamma_t \overset{\mathcal{D}^\dagger}{\to} \chi_1^2, \quad \begin{array}{ll} \text{under } H_0, & \text{for } m \leq t \leq T, \\ \text{under } H_{A,1}, & \text{for } m \leq t < \tau \text{ and } \tau + m - 1 \leq t \leq T, \\ \text{under } H_{A,2}, & \text{for } m \leq t < \tau, \end{array} \tag{18}$$

for almost all realisations of $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$ and $\{\xi_j(t), 1 \leq j \leq R,\ m \leq t \leq T\}$.
Under Assumptions 1-5, as $\min(N, m, R, W) \to \infty$*, it holds that*

$$\frac{1}{W}\Gamma_t \overset{P^\dagger}{\to} \frac{\int_{-\infty}^{+\infty} \left|I_{[0,\infty)}(u) - G_\psi(0)\right|^2 dF_\psi(u)}{G_\psi(0)[1 - G_\psi(0)]}, \quad \begin{array}{ll} \text{under } H_{A,1}, & \text{for } t_{N,m}^* \leq t < t_{N,m}^{**}, \\ \text{under } H_{A,2}, & \text{for } t_{N,m}^* \leq t \leq T, \end{array} \tag{19}$$

for almost all realisations of $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$ and $\{\xi_j(t),\ 1 \leq j \leq R,\ m \leq t \leq T\}$.

Theorem 2 is, again, an intermediate result. It states that Γ_t has (asymptotically) a chi-square distribution under the null of no breaks; further, by construction the sequence

$\{\Gamma_t\}_{t=m}^T$ is independent across t conditional on the sample. We now discuss how these two basic facts can be employed in order to propose a monitoring scheme for the on-line detection of breaks in the factor structure.

3.2. Sequential monitoring of factor models

We base our sequential monitoring procedure on the theory developed in Horváth *et al.* [37]. Recall that, after collecting m observations, we monitor our model over the period $m+1 \leq t \leq T$, which has size denoted as $T_m = T - m$. We then consider a monitoring procedure based on the detector

$$d(k;m) = \left| \sum_{t=m+1}^{m+k} \frac{\Gamma_t - 1}{\sqrt{2}} \right|, \quad 1 \leq k \leq T_m, \tag{20}$$

which covers the entire monitoring period. In other words our detector is made of the cumulative sum of the centered and standardized version of the sequence $\{\Gamma_t\}_{t=m}^T$, obtained by double randomisation. Other detectors, differing form (20) only with respect to the start of the monitoring period, could be also suggested. In particular, Kirch and Weber [42] suggest to use a rolling window, thus starting the monitoring procedure at $t = m+k-h+1$ for some $\underline{h} < h < k$, with $\underline{h}$ large enough. The asymptotic properties of such alternative detector can be derived in a way similar to the results proved in this section and therefore are not discussed in this paper. In light of Theorem 2, a break implies a shift in the mean of Γ_t and therefore in the detector (20). Therefore, our monitoring scheme looks for large deviations of $d(k;m)$ from its null-distribution.

Given the stopping rule

$$\widehat{k}_m = \begin{cases} \inf\{1 \leq k \leq T_m, \text{ such that } d(k;m) \geq \nu(k;m)\}, \\ T_m \text{ if the above does not hold in } 1 \leq k \leq T_m, \end{cases} \tag{21}$$

we define the estimated change-point location as $\widehat{\tau}_m = \widehat{k}_m + m$. The threshold function in (21) is defined as (see Horváth *et al.* [37] and Horváth *et al.* [38])

$$\nu(k;m) = c_{\alpha,m}\nu^*(k;m), \tag{22}$$

$$\nu^*(k;m) = m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\eta}, \ \eta \in \left[0, \frac{1}{2}\right], \tag{23}$$

where $c_{\alpha,m}$ is a critical value corresponding to a pre-specified level α. Depending on the choice of η, the critical value is defined as

$$P\left(\sup_{0\leq t\leq 1} \frac{|B(t)|}{t^{\eta}} \leq c_{\alpha,m}\right) = 1-\alpha, \quad \text{for } \eta \in \left[0, \frac{1}{2}\right), \tag{24}$$

where $\{B(t),\ 0 \leq t \leq 1\}$ denotes a standard Wiener process, or

$$c_{\alpha,m} = \frac{D_m - \ln[-\ln(1-\alpha)]}{A_m}, \quad \text{for } \eta = \frac{1}{2}, \tag{25}$$

with $A_m = (2\ln\ln m)^{1/2}$ and $D_m = 2\ln\ln m + \frac{1}{2}\ln\ln\ln m - \frac{1}{2}\ln\pi$. Note that in (24) $c_{\alpha,m}$ does not depend on m, whilst it does in (25). Note also that Chu *et al.* [23], albeit in a

different context, choose $\eta = 0$. It is well known that tests based on $\eta = 0$ have the smallest power, which on the contrary increases as η increases (see the discussion in Horváth *et al.* [37]).

In order to derive our main theorem, we also need the following assumptions.

Assumption 8. *It holds that (i)* $T_m = O\left(m^{\varkappa}\right)$ *for some* $\varkappa \geq 1$*; (ii)* $\liminf_{m\to\infty} \frac{T_m}{m} > 0$*; (iii)* $T_m > \tau + C_0 m^{1/2+\epsilon}$ *for* $\epsilon > 0$ *such that* $\frac{N^{1-\delta}}{m^{1/2-\epsilon}} \to C_1$.

Assumption 9. *It holds that (i)* $\int_{-\infty}^{+\infty} |u|^{4+2\delta}\, dF_{\psi}(u) < \infty$*;*

$$\textit{(ii)}\quad m^{1/2+\epsilon}\left\{W^{-1} + W\left[h\left(\frac{R}{\widetilde{l}\left(N,m,R\right)}\right)\right]^{-2} + \left[h\left(\frac{R}{\widetilde{l}\left(N,m,R\right)}\right)\right]^{-1}\right\} \to 0,$$

for some $\epsilon > 0$.

Assumption 8 is the same as equation (1.12) in Horváth *et al.* [38], and it essentially requires that the monitoring goes on for a sufficiently long time, longer than the initial training period m. In particular, we need to monitor for a number of periods of order at least $m^{1/2}$. Assumption 9 strengthens Assumption 6*(ii)*, and it is needed to prove a moment condition for the sequence $\{\Gamma_t\}_{t=m}^{T}$ which will enable a Central Limit Theory to hold. Our main result follows.

Theorem 3. *Let Assumptions 1-9 hold. Under* H_0 *it holds that, as* $\min\left(N,m,R,W\right) \to \infty$

$$P^{\dagger}\left(\max_{1\leq k\leq T_m} \frac{d\left(k;m\right)}{\nu^{*}\left(k;m\right)} \leq x\right) \to P\left(\sup_{0\leq t\leq 1} \frac{\left|B\left(t\right)\right|}{t^{\eta}} \leq x\right), \quad \textit{for } \eta \in \left[0, \frac{1}{2}\right), \tag{26}$$

$$P^{\dagger}\left(\max_{1\leq k\leq T_m} \frac{d\left(k;m\right)}{\nu^{*}\left(k;m\right)} \leq \frac{x + D_m}{A_m}\right) \to e^{-e^{-x}}, \quad \textit{for } \eta = \frac{1}{2}, \tag{27}$$

for almost all realisations of $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$ *and* $\{\xi_j(t),\ 1 \leq j \leq R,\ m \leq t \leq T\}$ *and for* $x \in \mathbb{R}^{+}$.
Under $H_{A,1}$ *and* $H_{A,2}$*, as* $\min\left(N,m,R,W\right) \to \infty$*, and for a given significance level* α*, it holds that*

$$c_{\alpha,m}^{-1} \max_{1\leq k\leq T_m} \frac{d\left(k;m\right)}{\nu^{*}\left(k;m\right)} \overset{P^{\dagger}}{\to} \infty, \quad \textit{for all } \eta \in \left[0, \frac{1}{2}\right], \tag{28}$$

for almost all realisations of $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$ *and* $\{\xi_j(t),\ 1 \leq j \leq R,\ m \leq t \leq T\}$*, where* $c_{\alpha,m}$ *is defined in* (24) *when* $\eta < \frac{1}{2}$ *and in* (25) *when* $\eta = \frac{1}{2}$.

The main implication of Theorem 3 is summarized in the following result (recall that $T = T_m - m$):

Corollary 1. *Under the assumptions of Theorem 3 it holds that, as* $\min\left(N,m,R,W\right) \to \infty$

$$P^{\dagger}\left(\widehat{\tau}_m < T\right) \leq \alpha, \quad \textit{under } H_0, \tag{29}$$

$$P^{\dagger}\left(t_{N,m}^{*} \leq \widehat{\tau}_m < t_{N,m}^{**}\right) = 1, \quad \textit{under } H_{A,1} \tag{30}$$

$$P^{\dagger}\left(t_{N,m}^{*} \leq \widehat{\tau}_m \leq T\right) = 1, \quad \textit{under } H_{A,2}, \tag{31}$$

for almost all realisations of $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$ *and* $\{\xi_j(t),\ 1 \leq j \leq R,\ m \leq t \leq T\}$.

The notion of size implied by (29), in this context, is very different from the one usually considered in the literature. The purpose of the procedure is to keep the false rejection probability as little as possible, and therefore (at a minimum) below the threshold α, rather than making it close to α. This makes the monitoring procedure different from the standard Neyman-Pearson paradigm (and, in general, from a multiple testing exercise): given that the monitoring horizon keeps expanding, the purpose of $c_{\alpha,m}$ is to ensure that the chance of a false break detection is as little as possible – see also similar comments in Horváth *et al.* [38].

3.3. Delay in change-point detection

A consequence of our approach is that monitoring for a structural change (despite being in a high-dimensional set-up) can be treated as in a classical time series framework. In particular, in addition to the consistency of the procedure, a natural question is how much would the delay be in detecting a break. In order to formally address this issue, one can directly use the results by Aue and Horváth [4]; hereafter, we provide a heuristic discussion of the magnitude of the delay within our setup.

Consider the notation $a_n = \Omega(b_n)$ to indicate that the magnitude of the sequence a_n is not smaller than that of b_n, viz. $a_n > Cb_n > 0$. Then, by construction, $\{\Gamma_t\}_{t=m}^T$ has, under the alternative, a "large" shift in the mean after $t^*_{N,m}$, where $t^*_{N,m}$ is such that (recall (14))

$$t^*_{N,m} - \tau = \Omega\left(\frac{m}{N^{1-\delta}}\right). \tag{32}$$

Defining β such that $N = m^\beta$, and using (12), it is possible to analyse (32) for various relative rates of divergence of m and N as they pass to infinity. When $\beta > \frac{1}{2}$, we have that $\delta = 1 - \frac{1}{2\beta} + \epsilon$ for an arbitrarily small value of ϵ. Thus, by (32)

$$t^*_{N,m} = \tau + \Omega\left(\frac{m}{m^{\beta(1-\delta)}}\right) = \tau + \Omega\left(m^{1/2+\epsilon'}\right),$$

where $\epsilon' > 0$ is arbitrarily small. Thus, when N is not much smaller than m, or even larger, the change-point is detected with a delay, $t^*_{N,m} - \tau$, which is of order at least $m^{1/2}$. By the same token, whenever $\beta \leq \frac{1}{2}$, i.e. N is much smaller than m, we have that $\delta = \epsilon$ for an arbitrarily small value of ϵ, so that

$$t^*_{N,m} = \tau + \Omega\left(m^{1-\beta(1-\delta)}\right),$$

and, by elementary arguments, it follows that $m^{1-\beta(1-\delta)} = \Omega(m^{1/2+\epsilon'})$: the delay, in this case, might be bigger. This is in line with the intuition that a break will cause Γ_t - and consequently the detector - to diverge as fast as N: the lower N, the lower the divergence rate, and the less effective the detecion of breaks. Finally, it is interesting to consider the ultra high-dimensional case, $N = \exp(m)$. By (12), it holds that $\delta = 1 - (1-\epsilon)\frac{\ln m}{2m}$ for an arbitrarily small value of ϵ. Hence, (32) yields

$$t^*_{N,m} = \tau + \Omega\left(\frac{m}{\exp\left((1-\delta)\, m\right)}\right) = \tau + \Omega\left(m^{1/2+\epsilon'}\right),$$

again. In essence, in all cases considered there is a delay in the detection of breaks which is greater than $C_0 m^{1/2}$, but smaller than $C_0 m$ - that is, rescaling the delay by the sample

size, this vanishes.

4. Applying the test under general circumstances

The purpose of this section is to discuss how the test could be applied under slightly different assumptions than the ones above, and up to which extent such assumptions can be relaxed. More substantive extensions, which involve modifications of the test, are briefly discussed in the concluding remarks in Section 7.

4.1. Weak factors and local alternatives

The theory developed in this paper - starting from Assumption 2 - implicitly requires that, when a new factor appears as a consequence of a break, this should be a pervasive factor. Indeed, part *(i)* of the assumption entails that spiked eigenvalues must diverge at a rate N, i.e. a "strong" factor model. However, the literature has also considered cases in which one or more common factor may be less pervasive, thus leading to a covariance matrix which has some eigenvalues passing to infinity at a rate N^{κ}, for $\kappa \in (0,1)$. A possible example of weak factors arises when considering jointly macroeconomic data of different countries: global factors are strong since they are likely to affect all countries; however national factors, although strong within a given country, will affect only a subset of all variables considered and can be seen as weak – see e.g. the empirical study in Moench *et al.* [50]. Estimation of factor models in the presence of such "weak" or "local" factors have been paid considerable attention by the literature - see De Mol *et al.* [27], Onatski [53], in the same setting as ours and, in a slightly different context, Lam and Yao [46]. The notion of weak factors is intertwined with that of a local alternative hypothesis where the break does happen but it is "small", for example when a break is caused by a change of only some, but not all, loadings. We focus on the (algebraically simpler) case of $H_{A,2}$. Consistently with the literature on weak factors, we allow the $(r+1)$-th eigenvalue to behave as

$$\lambda^{(r+1)}(t) = C_0 N^{\kappa}, \quad \text{for} \quad \tau \leq t \leq T, \tag{33}$$

for $\kappa \in (0,1)$, while it is bounded for all other values of t.

We now discuss heuristically under which conditions such small breaks can be detected; we consider for simplicity the case $\eta < \frac{1}{2}$. We know that, based on Theorem 2, a break in the $(r+1)$-th largest eigenvalue enters the sequence $\{\Gamma_t\}_{t=m}^{T}$ as a shift in its mean: this is essentially the way in which the monitoring procedure picks up the presence of a break. In particular, by analysing the proof of Theorem 1 and using a Mean Value argument, it follows that

$$\Theta_t \approx R \int_{-\infty}^{+\infty} \left| G_{\phi}\left(u\phi_{N,m}^{-1}(t)\right) - G_{\phi}(0) \right|^2 dF_{\phi}(u) \approx C_0 R \phi_{N,m}^{-2}(t), \tag{34}$$

for any $t \geq t_{N,m}^{*}$ for which $H_{A,2}$ holds. Then, from (19), for the same values of $t \geq t_{N,m}^{*}$ for which (34) holds we have

$$\Gamma_t \approx W \int_{-\infty}^{+\infty} \left| G_{\psi}\left(u\psi_{N,m,R}^{-1}(t)\right) - G_{\psi}(0) \right|^2 dF_{\psi}(u) \approx C_0 W \psi_{N,m,R}^{-2}(t). \tag{35}$$

Consider now the case where $g(\cdot)$ in (13) and $h(\cdot)$ in (17) are both the identity function. Recalling the notation $N = m^{\beta}$, and noting that by (33) we have $\phi_{N,m}(t) \approx N^{\kappa-\delta}$, by (17), (34), and (35), we have

$$\Gamma_t \approx C_0 W R^{-2} (\ln N)^{8+\epsilon} (\ln R)^{4+\epsilon} N^{4(\kappa-\delta)} = \Delta_{N,R,W}.$$

Upon inspecting the proof of Theorem 3, in order for the procedure to detect a break, it is required that $m^{1/2}\Delta_{N,R,W} \to \infty$, as $\min(m, N, R, W) \to \infty$. Therefore, if $\Delta_{N,R,W} \to \infty$ the break is always detectable. If instead $\Delta_{N,R,W} \to 0$, we are in presence of a shrinking break. By Assumption 7, a sufficient condition to have a shrinking break is

$$\kappa \leq \delta, \tag{36}$$

and a necessary condition for the break to be detectable is

$$\kappa > \delta - \frac{1}{8\beta}. \tag{37}$$

Consider first the case $N > m^{1/2}$. Then, by definition of δ we always have a shrinking break whenever $\kappa \leq 1 - \frac{1}{2\beta}$ and moreover a new weak factor is detected if at least $\kappa > 1 - \frac{5}{8\beta}$. This entails that we can hope to detect new weak factor for any $\kappa > 0$ only if $\beta < \frac{5}{8}$; conversely, for larger values of β the range of values of κ for which we can detect a new factor is reduced, e.g. for $N = m$, we must have at least $\kappa > \frac{3}{8}$.

Turning to the case $N \leq m^{1/2}$, since we can choose δ to be infinitesimally small, (36) is never satisfied but (37) is always satisfied and in general we cannot say more about the ability of our procedure to detect a shrinking break. However, we note that in the case the case $N = R = W$, as in Sections 5 and 6 below, a necessary and sufficient condition for a break to be shrinking and detectable is $\delta + \frac{1}{4} - \frac{1}{8\beta} < \kappa < \delta + \frac{1}{4}$, and when $N \leq m^{1/2}$ a new factor is always detected regardless of κ.

4.2. Heteroskedasticity in the idiosyncratic component

The main assumptions in the paper are spelt out with respect to $X_{i,t}$, avoiding to make any comments on the properties of $u_{i,t}$ across time. We now discuss the behaviour of the test in the presence of heteroskedasticity, which is not explicitly considered (although not ruled out) by Assumption 2. For the sake of simplicity, we consider the case of an abrupt change in the covariance matrix of $\{u_{i,t}\}_{i=1}^{N}$, although more general forms of heteroskedasticity could also be considered.

To illustrate this, we consider a simple example where the covariance matrix $E(u_t u_t')$ undergoes an abrupt change of size Δ_u after a point in time, say τ^*:

$$E\left(u_t u_t'\right) = \begin{cases} \Sigma_u & \text{for} \quad m \leq t < \tau^*, \\ \Sigma_u + \Delta_u & \quad\quad \tau^* \leq t \leq T, \end{cases}$$

where Δ_u affects some or even all covariances. The only condition we require in order for our test to be applicable is $\omega^{(1)}\left(m^{-1}\sum_{k=t-m+1}^{t} E\left(u_t u_t'\right)\right) \leq C_0$ for each $t \geq m$, where the notation $\omega^{(1)}(A)$ is understood to represent the largest eigenvalue of a matrix A. This holds, when $t < \tau^*$, as long as $\omega^{(1)}(\Sigma_u) \leq C_0$. When $t \geq \tau^*$, using Weyl's inequality it

follows that

$$\omega^{(1)}\left(\frac{1}{m}\sum_{k=t-m+1}^{t}E\left(u_t u_t'\right)\right)\leq\omega^{(1)}\left(\Sigma_u\right)+\omega^{(1)}\left(\Delta_u\right), \tag{38}$$

which is bounded as long as $\omega^{(1)}\left(\Sigma_u\right)\leq C_0$ and $\omega^{(1)}\left(\Delta_u\right)\leq C_1$. In essence, as long as the perturbation matrix Δ_u is not too big, and therefore as long as the changes in the covariance structure of the idiosyncratic are not too big, our test can still be applied.

Condition (38) has interesting implications. Consider a break such that $\Delta_u = diag\left\{d_i\right\}$, with $0\leq d_i\leq C_0$ for all $1\leq i\leq N$. In such a case, where the variances of the error terms all undergo a change (potentially), but the covariance structure does not change, it would hold that $\omega^{(1)}\left(\Delta_u\right)\leq C_0$: even a large (but of finite size) break in the variance of the idiosyncratic components does not alter the structure of the eigenvalues of $E\left(X_t X_t'\right)$, by introducing a spurious spiked eigenvalue. Thus, an interesting question about the robustness of our procedure is: when is a break in the idiosyncratic component strong enough to be confused with a break in the factor structure? By the same (heuristic) token as above, the eigenvalue structure of $E\left(X_t X_t'\right)$ would change if, for argument's sake, $\omega^{(1)}\left(m^{-1}\sum_{k=t-m+1}^{t}E\left(u_t u_t'\right)\right)=C_0 N^{\varepsilon}$ with $\varepsilon\in(0,1]$. By Weyl's inequality assuming for simplicity that there is no break in the factor component

$$\omega^{(1)}\left(\frac{1}{m}\sum_{k=t-m+1}^{t}E\left(u_t u_t'\right)\right)\geq\omega^{(N)}\left(\Sigma_u\right)+\omega^{(1)}\left(\Delta_u\right)\geq\omega^{(1)}\left(\Delta_u\right). \tag{39}$$

Therefore, a sufficient condition would be $\omega^{(1)}\left(\Delta_u\right)=C_0 N^{\varepsilon}$. Moreover, given that $\omega^{(1)}\left(\Delta_u\right)\geq N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\Delta_u\right\}_{i,j}$, then (39) suggests that a break which is "sufficiently pervasive", so that it affects not merely the variances of the idiosyncratic components, but also their covariances (without needing to be necessarily huge), could introduce a spiked eigenvalue in $E\left(X_t X_t'\right)$. In such cases our procedure might detect τ^* as a change-point even if the signal component does not change - see also the same phenomenon documented in the off-line case by Barigozzi *et al.* [12].

4.3. Extensions to consider further alternative hypotheses

So far, we have focused our attention onto two empirically relevant but very specific forms of alternative hypotheses: a possible change in the loadings - $H_{A,1}$ - and a possible increase in the number of factors - $H_{A,2}$. However, our methodology is sufficiently general to be adapted (with minor modifications) to other cases also. A leading example is the case in which $q\geq 1$ factors vanish, viz.

$$H_{A,3}:\begin{cases} X_{i,t}=\sum_{j=1}^{r}a_{ij}f_{jt}+u_{i,t} & \text{for} \quad 1\leq t<\tau \\ X_{i,t}=\sum_{j=1}^{r-q}\widetilde{a}_{ij}f_{jt}+u_{i,t} & \quad\quad\quad \tau\leq t\leq T \end{cases}. \tag{40}$$

Note that, in (40), we can entertain the possibility that the loadings of the non-vanishing factors may also be subject to changes, although this is not required. For simplicity consider the case $q=1$, then under (40), it can be noted that the r-th eigenvalue of the covariance matrix of $X_{i,t}$ is spiked before τ, and bounded thereafter. This suggests that testing for (40) can be based on $\widehat{\lambda}^{(r)}(t)$. Since under the null (in essence, on account of Lemma 2) $N^{-\delta}\widehat{\lambda}^{(r)}(t)\to\infty$, whereas under the alternative $N^{-\delta}\widehat{\lambda}^{(r)}(t)\to 0$, one round of randomisation is enough to have a sequence of test statistics which behaves like $\left\{\Gamma_t\right\}_{t=m}^{T}$ under the

null - that is, which (conditional on the sample) is *i.i.d.*, has moments that exist up to any order, and has an asymptotic chi-square distribution, with mean and variance that can be approximated with a polynomially vanishing error. Hence, monitoring can be again carried out as proposed in Section 3.

5. Monte Carlo simulations

Under H_0 we simulate data according to the stable factor model (4):

$$X_{i,t} = a_i' f_t + u_{i,t}, \quad 1 \leq i \leq N,\ 1 \leq t \leq T.$$

In particular, we fix $N = 100$, and we consider $r \in \{1, 2, 3, 4\}$ factors. As far as the time dimension is concerned, we consider burn-in periods and thus sample sizes of dimension $m \in \{50, 75, 100, 125, 150, 175, 200, 225, 250\}$. We monitor our model for 1000 periods (that is, we set $T = 1000$). We simulate each element of the loadings vector a_i as $i.i.d.\mathcal{N}(0, 1)$; we assume some time dependence in the common factors through a causal VAR(1) process

$$f_t = H f_{t-1} + e_t, \quad 1 \leq t \leq T,$$

where $e_t \sim i.i.d.\mathcal{N}(0, I_r)$ and the matrix H has maximum absolute value of the eigenvalues equal to 0.7. The $N \times T$ matrix of idiosyncratic components u is generated as $u = D\varepsilon G$, where the $NT \times 1$ vector of stacked columns of ε is $i.i.d.\mathcal{N}(0, I_{NT})$ and D and G are two $N \times N$ and $T \times T$ Toeplitz matrices with entries, in the k-th diagonal place, given by 0.3^{k-1} and 0.5^{k-1} respectively. Finally, we have set the signal-to-noise ratio to $\frac{Var(X_{i,t})}{Var(u_{i,t})} = 2$ for all $1 \leq i \leq N$.

Under the alternative, we consider breaks to occur at the change-point $\tau = 500$ under the two schemes:

$$X_{i,t} = a_i' f_t\ I[t < \tau] + \widetilde{a}_i' f_t\ I[t \geq \tau] + u_{i,t}, \quad 1 \leq i \leq N,\ 1 \leq t \leq T, \tag{41}$$

$$X_{i,t} = a_i' f_t + b_i g_t\ I[t \geq \tau] + u_{i,t}, \quad 1 \leq i \leq N,\ 1 \leq t \leq T. \tag{42}$$

In (41), we consider the case in which all loadings undergo a change, i.e. $H_{A,1}$; all the elements of a_i and $\widetilde{a}_i$ are generated as $i.i.d.\mathcal{N}(0, 1)$. Scheme (42) refers to a break owing to a new common factor, g_t, appearing, i.e. $H_{A,2}$; the loadings b_i are generated as $i.i.d.\mathcal{N}(0, 1)$, and we simulate g_t as the causal AR(1)

$$g_t = \varphi g_{t-1} + v_t, \quad 1 \leq t \leq T,$$

with $v_t \sim i.i.d.\mathcal{N}(0, 1)$ and $\varphi = 0.7$. The idiosyncratic components are generated as before.

All results of the test are computed when setting $\eta = 0.45$ and $\eta = 0.5$. The critical values used in the case $\eta = 0.45$ are taken from Horváth *et al.* [37]; in particular, when the significance level is $\alpha = 0.05$ the critical value is $c_{0.05} = 2.7992$ and when $\alpha = 0.1$ we have $c_{0.1} = 2.5437$. Regarding the double randomisation, we choose the functions $g(\cdot)$ in (13) and $h(\cdot)$ in (17) to be the identity, we set $W = R = N$, the distributions G_ϕ and G_ψ in steps $A1$ and $B1$ are chosen to be standard normals, while F_ϕ and F_ψ in steps $A2$ and $B2$ are chosen to have non-zero and equal mass at $\pm\sqrt{2}$.

In order to evaluate the performance of our procedure, we repeat simulations 500 times, and we consider a series of indicators.

Table 1: Empirical size - 5% and 10% significance

Fraction of detections in $[m+1, T]$

	$m=50$				$m=75$				$m=100$			
	$\eta=0.45$		$\eta=0.5$		$\eta=0.45$		$\eta=0.5$		$\eta=0.45$		$\eta=0.5$	
r	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	0.03	0.05	0.03	0.06	0.04	0.05	0.03	0.05	0.04	0.06	0.03	0.06
2	0.04	0.05	0.04	0.06	0.03	0.04	0.02	0.05	0.04	0.06	0.04	0.06
3	0.03	0.05	0.03	0.05	0.03	0.05	0.03	0.06	0.04	0.06	0.04	0.06
4	0.03	0.05	0.03	0.06	0.02	0.05	0.02	0.06	0.04	0.05	0.03	0.06
	$m=125$				$m=150$				$m=175$			
	$\eta=0.45$		$\eta=0.5$		$\eta=0.45$		$\eta=0.5$		$\eta=0.45$		$\eta=0.5$	
r	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	0.05	0.06	0.05	0.06	0.04	0.05	0.03	0.05	0.04	0.07	0.04	0.06
2	0.03	0.05	0.03	0.06	0.04	0.05	0.03	0.05	0.05	0.06	0.04	0.07
3	0.03	0.05	0.03	0.05	0.03	0.05	0.03	0.05	0.04	0.08	0.04	0.08
4	0.03	0.06	0.03	0.07	0.04	0.07	0.05	0.08	0.04	0.06	0.05	0.06
	$m=200$				$m=225$				$m=250$			
	$\eta=0.45$		$\eta=0.5$		$\eta=0.45$		$\eta=0.5$		$\eta=0.45$		$\eta=0.5$	
r	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	0.05	0.07	0.05	0.07	0.05	0.07	0.05	0.07	0.04	0.07	0.04	0.08
2	0.04	0.05	0.03	0.05	0.04	0.06	0.04	0.07	0.04	0.07	0.04	0.07
3	0.05	0.07	0.04	0.08	0.03	0.04	0.03	0.04	0.03	0.05	0.04	0.05
4	0.04	0.06	0.04	0.06	0.04	0.06	0.04	0.07	0.04	0.07	0.04	0.08

Table 2: Power - loadings change - 5% significance

Fraction of detections in $[\tau, \tau+m)$

$\eta=0.45$	m								
r	50	75	100	125	150	175	200	225	250
1	0.96	0.95	0.96	0.96	0.96	0.96	0.96	0.95	0.95
2	0.58	0.97	0.97	0.96	0.96	0.96	0.97	0.95	0.98
3	0.01	0.74	0.97	0.97	0.96	0.97	0.96	0.97	0.96
4	0.00	0.03	0.80	0.94	0.96	0.94	0.96	0.96	0.96
$\eta=0.5$	m								
r	50	75	100	125	150	175	200	225	250
1	0.96	0.95	0.96	0.96	0.97	0.95	0.96	0.96	0.96
2	0.44	0.97	0.97	0.96	0.96	0.96	0.97	0.95	0.98
3	0.00	0.62	0.97	0.97	0.96	0.96	0.97	0.96	0.97
4	0.00	0.01	0.65	0.95	0.97	0.95	0.96	0.96	0.97

(1) In Table 1 we report the fraction of false rejections over the whole monitoring period ($m+1 \leq t \leq T$), when no break is present, i.e. under H_0, and when testing at 5% and 10% significance levels. As expected the empirical size is always below the significance level.

(2) In Tables 2, 3, 4 and 5 we show the fraction of detections for which $\tau \leq \widehat{\tau}_m < \tau+m-1$, when a break takes place under $H_{A,1}$ or $H_{A,2}$ and when testing at 5% and 10%

Table 3: Power - loadings change - 10% significance

Fraction of detections in $[\tau, \tau + m)$

$\eta = 0.45$					m				
r	50	75	100	125	150	175	200	225	250
1	0.94	0.93	0.95	0.93	0.94	0.93	0.94	0.94	0.93
2	0.66	0.95	0.95	0.93	0.95	0.93	0.94	0.94	0.95
3	0.01	0.81	0.95	0.95	0.94	0.95	0.95	0.94	0.94
4	0.00	0.06	0.87	0.94	0.95	0.92	0.93	0.94	0.94
$\eta = 0.5$					m				
r	50	75	100	125	150	175	200	225	250
1	0.94	0.93	0.95	0.92	0.93	0.93	0.93	0.93	0.93
2	0.59	0.94	0.95	0.92	0.94	0.93	0.94	0.93	0.95
3	0.01	0.73	0.94	0.95	0.94	0.95	0.94	0.93	0.94
4	0.00	0.03	0.80	0.93	0.94	0.91	0.93	0.94	0.93

Table 4: Power - new factor appears - 5% significance

Fraction of detections in $[\tau, \tau + m)$

$\eta = 0.45$					m				
r	50	75	100	125	150	175	200	225	250
1	0.93	0.93	0.92	0.95	0.92	0.95	0.93	0.93	0.92
2	0.78	0.96	0.97	0.95	0.95	0.96	0.96	0.95	0.96
3	0.10	0.89	0.97	0.98	0.96	0.97	0.95	0.95	0.97
4	0.00	0.27	0.91	0.96	0.96	0.96	0.95	0.95	0.96
$\eta = 0.5$					m				
r	50	75	100	125	150	175	200	225	250
1	0.95	0.94	0.93	0.96	0.95	0.96	0.95	0.93	0.94
2	0.71	0.96	0.97	0.96	0.95	0.96	0.96	0.96	0.97
3	0.06	0.85	0.97	0.98	0.96	0.98	0.96	0.96	0.98
4	0.00	0.16	0.89	0.96	0.96	0.95	0.95	0.96	0.97

significance levels, setting either $\eta = 0.45$ or $\eta = 0.5$. Results show that the test does have power versus the two alternative hypotheses considered in this paper. As the construction of the test and the theory would suggest, the power declines as r, the original, pre-break number of factors, increases: in essence, the test checks whether an eigenvalue is large, and the magnitude of the $(r+1)$-th largest eigenvalue declines with r. Still, even when $r = 4$, the test has high power when $m \geq 100$ in all cases considered, and, in presence of a new factor appearing (see Tables 4 and 5), even when $m \geq 50$. An interesting feature of the test is the case $\eta = 0.5$: although in theory this choice yields the highest power, it is well known that convergence to the extreme value distribution is very slow, leading to larger than correct critical values, and, consequently, to lower power (see the comments in Csörgő and Horváth [26]). However, considering the discrepancy between the power when $\eta = 0.45$ and $\eta = 0.5$, this is not always the case: tests based on the choice $\eta = 0.5$ have roughly the same power as for the case $\eta = 0.45$ whenever there is a change in the loadings, and also when there is a new factor appearing (at least for a sample size $m \geq 100$).

Table 5: Power - new factor appears - 10% significance

Fraction of detections in $[\tau, \tau + m)$

$\eta = 0.45$					m				
r	50	75	100	125	150	175	200	225	250
1	0.89	0.88	0.88	0.90	0.88	0.91	0.92	0.88	0.88
2	0.81	0.94	0.95	0.93	0.91	0.93	0.93	0.93	0.93
3	0.14	0.88	0.94	0.95	0.94	0.95	0.92	0.93	0.95
4	0.00	0.36	0.92	0.94	0.94	0.93	0.93	0.93	0.94
$\eta = 0.5$					m				
r	50	75	100	125	150	175	200	225	250
1	0.90	0.89	0.88	0.91	0.89	0.92	0.91	0.89	0.89
2	0.76	0.94	0.95	0.92	0.92	0.93	0.93	0.93	0.93
3	0.10	0.86	0.94	0.96	0.94	0.95	0.93	0.93	0.94
4	0.00	0.27	0.89	0.94	0.93	0.93	0.92	0.92	0.94

Last, notice that when considering $H_{A,2}$ (a new factor appearing), then we could also detect a change-point when $\tau + m - 1 \leq t \leq T$, but we do not report results in this case since power can only increase with respect to what shown in Tables 4 and 5.

(3) In Tables 6 and 7 we report the minimum, maximum, the 25th, 50th and 75th percentiles of the distribution of the estimated change-point locations, whenever under $H_{A,1}$ or $H_{A,2}$ a break is detected at $\widehat{\tau}_m$ such that $\tau \leq \widehat{\tau}_m \leq T$ and when testing at 10% significance levels. Given the results in Tables 3 and 5 we report those statistics only for $m = 100, 175, 250$. It is evident that the test detects a break with a delay which increases as r increases - this is in line with the comments in Section 4.1, since, as r grows, the r-th eigenvalue becomes smaller and smaller, thus being closer to a weak factor. Interestingly, there are virtually no differences between the cases of $\eta = 0.45$ and $\eta = 0.5$; similarly, different values of m also do not seem to alter results. Note that, as expected, the minimum values of the distribution of the estimated locations are, roughly speaking, of order $m^{1/2}$ all across the table.

6. An application to US industrial production data

We conclude with an application to a panel of US industrial production indexes. Specifically, we consider monthly growth rates for $N = 224$ sectorial indices, over the period from January 1972 to November 2015, for a total of $T = 527$ observations. Estimation is based on a sample of size $m = 60$, i.e. 5 years. Analysis of the whole dataset using rolling samples of size m suggests between one and two factors throughout - this result consistently follows using different procedures - namely, Trapani [62] testing procedure and the criteria by Bai and Ng [8] and Alessi *et al.* [2]. Therefore, we run our sequential testing procedure monitoring the first four factors, thus accounting both for at most two new factors emerging and for a change in all loadings. The test is run at 5% significance level and setting $\eta = 0.5$, hence using the critical values in (25).

The monitoring is implemented as follows. We begin at $t = m+1$; once the first change-point is detected at $\widehat{\tau}_1 \geq m+1$, we restart the estimation at $t = \widehat{\tau}_1$ and after m periods we restart monitoring at $t = \widehat{\tau}_1 + m + 1$. In general, given an estimated change-point $\widehat{\tau}_j$, with

Table 6: Location distribution - loadings change
(true change-point at $\tau = 500$)

		$\eta = 0.45$					$\eta = 0.5$				
m	r	min	25th	50th	75th	max	min	25th	50th	75th	max
100	1	504	516	520	526	551	505	516	521	526	551
	2	507	523	528	533	553	507	524	529	534	554
	3	509	529	536	543	573	509	531	538	545	586
	4	520	548	559	571	$> T$	520	553	563	578	$> T$
175	1	504	516	522	528	550	505	517	523	529	550
	2	508	523	529	535	553	509	524	530	537	555
	3	507	531	537	543	564	507	532	538	544	566
	4	514	536	544	551	580	514	538	546	553	591
250	1	502	518	523	529	551	505	519	524	529	553
	2	508	525	531	537	558	508	526	532	538	565
	3	508	531	538	547	568	508	533	540	548	570
	4	516	538	545	554	581	516	539	547	555	584

Table 7: Location distribution - new factor appears
(true change-point at $\tau = 500$)

		$\eta = 0.45$					$\eta = 0.5$				
m	r	min	25th	50th	75th	max	min	25th	50th	75th	max
100	1	502	514	519	525	547	505	515	520	525	548
	2	505	522	530	537	574	506	523	531	538	576
	3	513	532	541	550	598	514	534	542	552	598
	4	519	546	557	569	648	520	548	560	574	669
175	1	501	515	520	525	546	502	516	521	526	552
	2	507	524	531	537	564	507	524	532	539	572
	3	508	532	541	550	578	508	534	542	552	579
	4	518	543	552	562	598	518	544	555	565	599
250	1	502	514	520	527	552	502	514	521	527	553
	2	508	524	531	540	571	508	525	532	541	571
	3	510	533	542	551	590	510	534	543	552	589
	4	515	542	552	562	600	515	544	553	564	603

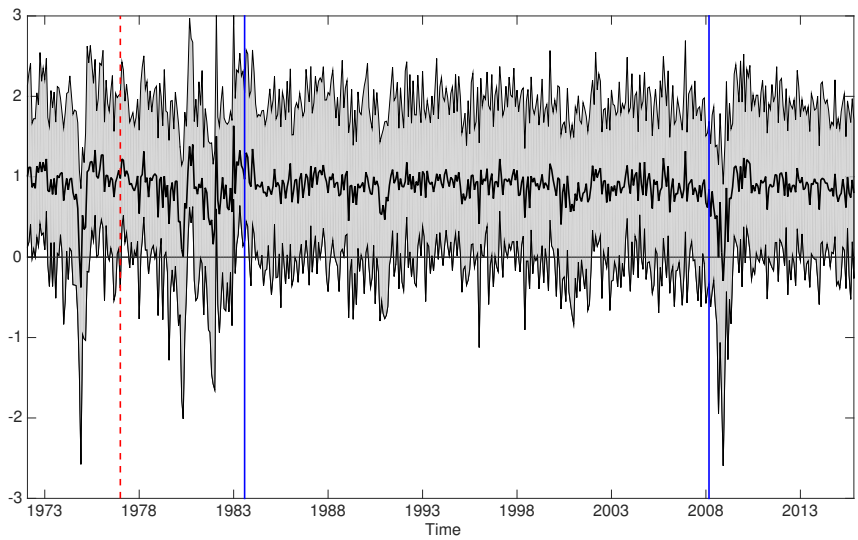
$j \geq 1$, we restart monitoring by computing the detector defined in (20) which in this case is defined as

$$d(k;m) = \left| \sum_{t=\widehat{\tau}_j+m+1}^{m+k} \frac{\Gamma_t - 1}{\sqrt{2}} \right|, \quad \widehat{\tau}_j + 1 \leq k \leq T - m.$$

Therefore, the monitoring window after the j-th change-point is of size $T - \widehat{\tau}_j - m$ and the estimated change-points $\widehat{\tau}_j$ are such that $\widehat{\tau}_{j+1} - \widehat{\tau}_j \geq m + 1$. We keep restarting the procedure as long as we have a monitoring window of non-zero length, that is as long as $T - \widehat{\tau}_j > m$; this allows the possibility for the last change-point to be detected in the interval $T - m \leq t \leq T$.

We find evidence of two change-points dated: (i) $\widehat{\tau}_1$: August, 1983; and (ii) $\widehat{\tau}_2$: March, 2008. The estimated locations are also shown in Figure 1 together with the joint panel of

Figure 1: Estimated change-point locations for US industrial production indexes



Black line: cross-sectional median of the data (monthly growth rates on yearly basis); blue lines: estimated change-point locations; red line: first period used for testing.

data. The first estimated change-point ($\widehat{\tau}_1$) clearly mark the start of the Great Moderation, i.e. a period of decrease in volatility of output and inflation, while the second one ($\widehat{\tau}_2$) takes place at the start of the US recession marked by the Great Financial Crisis. Last we discuss the delay of the estimated change-points. Concerning $\widehat{\tau}_1$, there is a general consensus that the start of the Great Moderation is to be dated in 1983, however a precise date is not available, see for example Stock and Watson [59]. We note here that if we consider the start of the Great Moderation to coincide with the end of the recession of the early 1980s, then the National Bureau of Economic Research (NBER) dates the start of the expansion of the US business cycle in December, 1982, thus the first change-point is detected with a delay of 4 time-periods. Concerning $\widehat{\tau}_2$ the NBER dates the start of the recession in December, 2007, therefore we detect the change-point with a delay of 3 time-points [3].

7. Conclusions

In this paper we develop a a family of monitoring procedures to detect a break in the signal component of a large factor model; to the best of our knowledge, this is the first contribution in high-dimensional factor models which proposes a sequential monitoring and testing procedure, as opposed to the extant literature where ex-post detection of breaks is usually considered. Our statistics are based on a well-known property of the $(r+1)$-th eigenvalue of the sample covariance matrix of the data: whilst under the null the $(r+1)$-th eigenvalue is bounded, under the alternative of a break (either in the loadings, or in the number of factors itself) it becomes spiked. Given that the sample eigenvalue does not have

[3] See `https://www.nber.org/cycles/US_Business_Cycle_Expansions_and_Contractions_20120423.pdf`

a known limiting distribution under the null, we regularise the problem by (doubly) randomising the test statistic in conjunction with sample conditioning, obtaining a sequence of *i.i.d.*, asymptotically chi-square statistics which are then employed to build the monitoring scheme. Numerical evidence shows that our procedure works very well in finite samples, with a very small probability of false detections and tight detection times in presence of a genuine change-point.

Building on the methodology proposed in this paper, there are at least two possible extensions which could be considered. Firstly, the results and methodology in this paper could be also used in the context of a non-stationary factor model, similar to the one considered in Bai [7], where the factors are allowed to have unit roots. In such case, the key theoretical result would be to show that in presence of r factors the first r eigenvalues of the matrix $m^{-2}\sum_{t=1}^{m} X_t X_t'$ diverge to positive infinity almost surely at some rate, whereas the remaining factors are a.s. bounded. Secondly, it is possible to extend the theory developed in this paper to the context of the generalised dynamic factor model by Forni *et al.* [30] or the factor model by Lam and Yao [46], which are based on the asymptotic behavior of the eigenvalues of the spectral density or the long-run covariance matrices, respectively. By studying the asymptotic behavior of the estimated eigenvalues of those matrices an appropriate test statistic based on these can be built. These, and other, extensions are under current investigations by the authors.

References

[1] Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, **81**, 1203–1227.

[2] Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate static factor models. *Statistics and Probability Letters*, **80**, 1806–1813.

[3] Andreou, E. and Ghysels, E. (2006). Monitoring disruptions in financial markets. *Journal of Econometrics*, **135**, 77–124.

[4] Aue, A. and Horváth, L. (2004). Delay time in sequential detection of change. *Statistics & Probability Letters*, **67**(3), 221–231.

[5] Aue, A., Hörmann, S., Horváth, L., Hušková, M., and Steinebach, J. G. (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory*, **28**, 804–837.

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

[7] Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, **122**, 137–183.

[8] Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70**, 191–221.

[9] Bai, Z. and Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, **106**, 167–177.

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

[11] Bandi, F. and Corradi, V. (2014). Nonparametric nonstationarity tests. *Econometric Theory*, **30**, 127–149.

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

[13] Bates, B. J., Plagborg-Møller, M., Stock, J. H., and Watson, M. W. (2013). Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics*, **177**, 289–304.

[14] Benaych-Georges, F. and Nadakuditi, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, **227**(1), 494–521.

[15] Benaych-Georges, F. and Nadakuditi, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, **111**, 120–135.

[16] Breitung, J. and Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, **163**, 71–84.

[17] Brodsky, B. (2010). Sequential detection and estimation of change-points. *Sequential Analysis*, **29**, 217–233.

[18] Cai, G.-H. (2006). Chover-type Laws of the Iterated Logarithm for weighted sums of ρ^*-mixing sequences. *International Journal of Stochastic Analysis*, **2006**.

[19] Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, **51**, 1281–1304.

[20] Chen, L., Dolado, J. J., and Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, **180**, 30–48.

[21] Chen, L. H. Y. (1978). A short note on the conditional Borel-Cantelli lemma. *The Annals of Probability*, **6**, 699–700.

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

[23] Chu, C., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, **64**, 1045–1066.

[24] Corradi, V. and 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.

[25] Corradi, V. and Swanson, N. R. (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics*, **182**, 100–118.

[26] Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*, volume 18. John Wiley & Sons Inc.

[27] De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, **146**, 318–328.

[28] El Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, **35**, 663–714.

[29] Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B*, **75**, 603–680.

[30] Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic Factor Model: Identification and estimation. *The Review of Economics and Statistics*, **82**, 540–554.

[31] Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models versus structural VARs. *Econometric Theory*, **25**, 1319–1347.

[32] Groen, J. J., Kapetanios, G., and Price, S. (2013). Multivariate methods for monitoring structural change. *Journal of Applied Econometrics*, **28**, 250–274.

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

[34] Hannan, E. J. (1970). *Multiple Time Series*. John Wiley & Sons.

[35] Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press.

[36] Horváth, L. and Trapani, L. (2019). Testing for randomness in a random coefficient autoregression model. *Journal of Econometrics*, **209**(2), 338–352.

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

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

[39] Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, **104**, 682–693.

[40] Jung, S. and Marron, J. S. (2009). PCA consistency in high dimension, low sample size context. *The Annals of Statistics*, **37**, 4104–4130.

[41] Kirch, C. and Tadjuidje Kamgaing, J. (2015). On the use of estimating functions in monitoring time series for change points. *Journal of Statistical Planning and Inference*, **161**, 25–49.

[42] Kirch, C. and Weber, S. (2017). Modified sequential change point procedures based on estimating functions. Technical report.

[43] Komlós, J., Major, P., and 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.

[44] Komlós, J., Major, P., and 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.

[45] Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society: Series B*, **57**, 613–658.

[46] Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, **40**, 694–726.

[47] Li, J., Todorov, V., Tauchen, G., and Lin, H. (2017). Rank tests at jump events. *Journal of Business & Economic Statistics*. forthcoming.

[48] Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics*, **197**, 101–129.

[49] Merikoski, J. K. and Kumar, R. (2004). Inequalities for spreads of matrix sums and products. *Applied Mathematics E-Notes*, **4**, 150–159.

[50] Moench, E., Ng, S., and Potter, S. (2013). Dynamic hierarchical factor models. *The Review of Economics and Statistics*, **95**, 1811–1817.

[51] Moricz, F. (1983). A general moment inequality for the maximum of the rectangular partial sums of multiple series. *Acta Mathematica Hungarica*, **41**, 337–346.

[52] Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, **92**, 1004–1016.

[53] Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, **168**, 244–258.

[54] Onatski, A., Moreira, M. J., and Hallin, M. (2014). Signal detection in high dimension: The multispiked case. *The Annals of Statistics*, **42**, 225–254.

[55] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, **17**, 1617–1642.

[56] Pearson, E. (1950). On questions raised by the combination of tests based on discontinuous distributions. *Biometrika*, **37**, 383–398.

[57] Sen, P. K. (1981). *Sequential nonparametrics: invariance principles and statistical inference*. Probability and mathematical statistics. Wiley New York.

[58] Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, **97**, 1167–1179.

[59] Stock, J. H. and Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, **39**, 3–33.

[60] Stock, J. H. and Watson, M. W. (2009). Forecasting in dynamic factor models subject to structural instability. In J. Castle and N. Shephard, editors, *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Oxford University Press.

[61] Stock, J. H. and Watson, M. W. (2011). Dynamic factor models. In M. Clements and D. Hendry, editors, *Oxford Handbook of Economic Forecasting*. Oxford University Press.

[62] Trapani, L. (2018). A randomized sequential procedure to determine the number of factors. *Journal of the American Statistical Association*, **113**(523), 1341–1349.

[63] Wang, W. and Fan, J. (2017). Asymptotics of empirical eigen-structure for high dimensional spiked covariance. *The Annals of Statistics*. forthcoming.

[64] Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 14150–14154.

[65] Yamamoto, Y. and Tanaka, S. (2015). Testing for factor loading structural change under common breaks. *Journal of Econometrics*, **189**, 187–206.

Appendix A. Technical appendix

Appendix A.1. Preliminary lemmas

This section contains technical results which are useful to prove the main theorems in the paper. Throughout this and the next section, E^* denotes expectation calculated with respect to P^*; similarly, $E^\dagger$ and $V^\dagger$ denote expectation and variance calculated with respect to $P^\dagger$. Also, whenever possible, we omit for ease of notation the dependence of $\xi_j(t)$, and of related quantities, on t.

Lemma A.1. *Let*

$$\overline{\lambda}_N(t) = \frac{1}{N}\sum_{p=1}^{N}\lambda^{(p)}(t) \quad \textit{and} \quad \widehat{\overline{\lambda}}_N(t) = \frac{1}{N}\sum_{p=1}^{N}\widehat{\lambda}^{(p)}(t).$$

Under Assumptions 1 and 2, it holds that

$$\begin{cases} \limsup_{N\to\infty}\overline{\lambda}_N(t) = \overline{\lambda}^{\sup}(t) < \infty \\ \liminf_{N\to\infty}\overline{\lambda}_N(t) = \overline{\lambda}^{\inf}(t) > 0 \end{cases},$$

for all $m \le t \le T$. *Furthermore, under Assumptions 1-3, it holds that, as* $\min\{N,m\}\to\infty$

$$\begin{cases} \limsup_{N\to\infty}\widehat{\overline{\lambda}}_N(t) = \overline{\lambda}^{\sup}(t) < \infty \\ \liminf_{N\to\infty}\widehat{\overline{\lambda}}_N(t) = \overline{\lambda}^{\inf}(t) > 0 \end{cases}.$$

Proof. See Lemmas 2.1 and A.1 in Trapani [62]. □

Lemma A.2. *Under Assumptions 1-4, it holds that*

$$\lim \sup_{N,m,R\to\infty}\frac{\Theta_t}{\widetilde{l}(N,m,R)} = 0 \quad a.s., \begin{array}{ll} \textit{under } H_{A,1}, & \textit{for } t^*_{N,m} \le t < t^{**}_{N,m}, \\ \textit{under } H_{A,2}, & \textit{for } t^*_{N,m} \le t \le T, \end{array} \tag{A.1}$$

and

$$\frac{\widetilde{l}(N,m,R)}{R}\times\frac{\Theta_t}{\widetilde{l}(N,m,R)} \to C > 0 \quad a.s., \begin{array}{ll} \textit{under } H_0, & \textit{for } m \le t \le T, \\ \textit{under } H_{A,1}, & \textit{for } m \le t < \tau, \\ & \textit{and } \tau + m - 1 \le t \le T, \\ \textit{under } H_{A,2}, & \textit{for } m \le t < \tau, \end{array} \tag{A.2}$$

as $\min(N,m,R)\to\infty$.

Proof. We begin with (A.1). For any $1 \le n \le N$, $1 \le s \le m$, $1 \le r \le R$, define

$$U_{n,s,r} = \int_{-\infty}^{+\infty}\left| r^{-1/2}\sum_{j=1}^{r}\left\{I\left[\xi_j \le u\phi^{-1}_{n,s}(t)\right] - G_\phi(0)\right\}\right|^2 dF_\phi(u).$$

We begin by showing that

$$\sum_{N=1}^{\infty}\sum_{m=1}^{\infty}\sum_{R=1}^{\infty}\frac{1}{mNR}P^*\left[\max_{1\le n\le N, 1\le s\le m, 1\le r\le R} U_{n,s,r} > \epsilon\widetilde{l}(N,m,R)\right] < \infty, \tag{A.3}$$

for any $\epsilon > 0$. Using the short-hand notation $\max_{n,s,r}$ for $\max_{1\leq n\leq N, 1\leq s\leq m, 1\leq r\leq R}$, Markov inequality implies that (A.3) follows if

$$\sum_{N=1}^{\infty}\sum_{m=1}^{\infty}\sum_{R=1}^{\infty}\frac{1}{mNR\widetilde{l}(N,m,R)}E^{*}\left|\max_{n,s,r}U_{n,s,r}\right| < \infty. \tag{A.4}$$

The maximal inequality contained in Theorem 2 in Moricz [51] entails that

$$E^{*}\left|\max_{n,s,r}U_{n,s,r}\right| \leq C_0 E^{*}\left|U_{N,m,R}\right|(\ln N)(\ln m)(\ln R). \tag{A.5}$$

Further, combining (A.25) with (A.26)-(A.27), it is easy to see that

$$E^{*}\left|U_{N,m,R}\right| \leq C_0 + C_1 R\phi_{N,m}^{-2}(t),$$

which holds under $H_{A,1}$ for $t_{N,m}^{*} \leq t < t_{N,m}^{**}$ and under $H_{A,2}$ for $t_{N,m}^{*} \leq t \leq T$. By Assumption 5, it holds that $E^{*}\left|U_{N,m,R}\right|$ is bounded. Then (A.3) follows immediately from (A.4).

Note now that for every triple (N,m,R), there is a triple of positive integers (k_1,k_2,k_3) such that $2^{k_1} \leq N < 2^{k_1+1}$, $2^{k_2} \leq m < 2^{k_2+1}$, $2^{k_3} \leq R < 2^{k_3+1}$. Further, there is also a triple of real numbers defined over $[0,1)$, say (ρ_1,ρ_2,ρ_3), such that $N = 2^{k_1+\rho_1}$, etc... Define now the short-hand notation

$$A_{k_1,k_2,k_3} = \left\{\omega : \max_{1\leq k_1\leq 2^{k_1+\rho_1}, 1\leq k_2\leq 2^{k_2+\rho_2}, 1\leq k_3\leq 2^{k_3+\rho_3}}\left|U_{k_1,k_2,k_3}\right| > \epsilon\widetilde{l}(k_1,k_2,k_3)\right\}.$$

By (A.3), it holds that

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty}\frac{2^{k_1+1}2^{k_2+1}2^{k_3+1}}{(2^{k_1+1}-1)(2^{k_2+1}-1)(2^{k_3+1}-1)}P^{*}(A_{k_1,k_2,k_3}) < \infty;$$

thus

$$\begin{aligned}&\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty}P^{*}(A_{k_1,k_2,k_3}) \leq \\ \leq\ & 2^3\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty}\frac{2^{k_1}2^{k_2}2^{k_3}}{(2^{k_1+1}-1)(2^{k_2+1}-1)(2^{k_3+1}-1)}P^{*}(A_{k_1,k_2,k_3}) < \infty.\end{aligned} \tag{A.6}$$

This result entails that $P^{*}(A_{k_1,k_2,k_3}\text{ i.o.}) = 1$, which is a *conditional* result. Let now X_{k_1,k_2,k_3} be the indicator of A_{k_1,k_2,k_3}, and note that A_{k_1,k_2,k_3} is conditional on the σ-field $\mathcal{F}_{k_1,k_2,k_3} = \{X_{i,t},\ 1\leq i\leq N, 1\leq t\leq m\}\cup\{\xi_j, 1\leq j\leq R\}$, which is non decreasing. Equation (A.6) implies that

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty}E(\mathrm{X}_{k_1,k_2,k_3}|\mathcal{F}_{k_1,k_2,k_3}) < \infty;$$

hence, by Theorem 1 in Chen [21], it holds that

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty} \mathrm{X}_{k_1,k_2,k_3} < \infty \quad \text{a.s.} \tag{A.7}$$

We note that the result by Chen [21] is for a series indexed by a single index, but his arguments can be readily generalised to the case of multi-index series. Equation (A.7) can be equivalently rewritten as

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty} P\left(A_{k_1,k_2,k_3}\right) < \infty, \tag{A.8}$$

which is an *unconditional* result. From (A.8), it is easy to see that

$$\frac{\max_{k_1,k_2,k_3}\left|U_{k_1,k_2,k_3}\right|}{\widetilde{l}\left(k_1,k_2,k_3\right)} \to 0 \text{ a.s.;}$$

this entails that

$$\frac{\left|U_{N,m,R}\right|}{\widetilde{l}\left(N,m,R\right)} \leq \frac{\max_{k_1,k_2,k_3}\left|U_{k_1,k_2,k_3}\right|}{\widetilde{l}\left(k_1,k_2,k_3\right)}\frac{\widetilde{l}\left(k_1,k_2,k_3\right)}{\widetilde{l}\left(N,m,R\right)} \leq \frac{\max_{k_1,k_2,k_3}\left|U_{k_1,k_2,k_3}\right|}{\widetilde{l}\left(k_1,k_2,k_3\right)} \to 0 \text{ a.s.,}$$

so that finally

$$\lim \sup_{N,m,R\to\infty} \frac{\left|U_{N,m,R}\right|}{\widetilde{l}\left(N,m,R\right)} = 0 \text{ a.s.,}$$

from which (A.1) follows.

Consider now (A.2). Under H_0, Lemma 1 entails that

$$P\left\{\omega : \lim_{N,m\to\infty} \phi_{N,m}\left(t\right) = 0\right\} = 1, \quad m \leq t \leq T,$$

so that we can assume henceforth that $\lim_{N,m\to\infty} \phi_{N,m}\left(t\right) = 0$ for $m \leq t \leq T$. Similarly this holds also under $H_{A,1}$ for $m \leq t < \tau$ and $\tau + m - 1 \leq t \leq T$ and under $H_{A,2}$ for $m \leq t < \tau$.

Also, by definition it holds that $E^{*}I\left[\xi_j \leq u\phi_{N,m}^{-1}(t)\right] = G_\phi\left(u\phi_{N,m}^{-1}(t)\right)$. Therefore

$$
\begin{aligned}
& G_\phi(0)\left[1-G_\phi(0)\right]\Theta_t = \\
&= \int_{-\infty}^{+\infty}\left|R^{-1/2}\sum_{j=1}^{R}\left\{I\left[\xi_j \leq u\phi_{N,m}^{-1}(t)\right] - G_\phi\left(u\phi_{N,m}^{-1}(t)\right) + G_\phi\left(u\phi_{N,m}^{-1}(t)\right) - G_\phi(0)\right\}\right|^2 dF_\phi(u) = \\
&= \int_{-\infty}^{+\infty}\left|R^{-1/2}\sum_{j=1}^{R}\left\{I\left[\xi_j \leq u\phi_{N,m}^{-1}(t)\right] - G_\phi\left(u\phi_{N,m}^{-1}(t)\right)\right\}\right|^2 dF_\phi(u) + \\
&\quad + R^{1/2}\int_{-\infty}^{+\infty}\left|G_\phi\left(u\phi_{N,m}^{-1}(t)\right) - G_\phi(0)\right|^2 dF_\phi(u) + \\
&\quad + 2\int_{-\infty}^{+\infty} R^{-1/2}\sum_{j=1}^{R}\left\{I\left[\xi_j \leq u\phi_{N,m}^{-1}(t)\right] - G_\phi\left(u\phi_{N,m}^{-1}(t)\right)\right\}\left[G_\phi\left(u\phi_{N,m}^{-1}(t)\right) - G_\phi(0)\right] dF_\phi(u).
\end{aligned}
$$

Note that, by (A.2), $G_\phi\left(u\phi_{N,m}^{-1}(t)\right) - G_\phi(0) = I_{[0,\infty)}(u) - G_\phi(0)$ as $N, m \to \infty$. Also, using similar arguments as in the proof of (A.1), it is easy to see that

$$
\limsup_{N,m,R\to\infty} \frac{\int_{-\infty}^{+\infty}\left|R^{-1/2}\sum_{j=1}^{R}\left\{I\left[\xi_j \leq u\phi_{N,m}^{-1}(t)\right] - G_\phi\left(u\phi_{N,m}^{-1}(t)\right)\right\}\right|^2 dF_\phi(u)}{\widetilde{l}(N,m,R)} = 0 \text{ a.s.;}. \tag{A.9}
$$

Equation (A.2) follows directly from (A.2), (A.9) (and the Cauchy-Schwartz inequality), and Assumption 4*(iii)*. We point out that the passages above follow closely to the proof of Theorem 4 in Horváth and Trapani [36]. □

Lemma A.3. *Under Assumptions 1-6 and 9(ii), it holds that, under H_0*

$$
\max_{1\leq k\leq T_m}\sqrt{\frac{m}{k(k+m)}}\left|\sum_{t=m+1}^{m+k}\left(E^{\dagger}(\Gamma_t) - 1\right)\right| = O\left(m^{-\epsilon}\right), \tag{A.10}
$$

$$
\max_{1\leq k\leq T_m}\sqrt{\frac{m}{k(k+m)}}\left|\sum_{t=m+1}^{m+k}\left(V^{\dagger}(\Gamma_t) - 2\right)\right| = O\left(m^{-\epsilon}\right), \tag{A.11}
$$

for some $\epsilon > 0$. Also

$$
E^{\dagger}\left|\Gamma_t\right|^{2+\delta} < \infty, \tag{A.12}
$$

for some $\delta > 0$.

Proof. We start with equation (A.10). By construction

$$\frac{E^{\dagger}\left(\Gamma_t\right)}{G_{\psi}\left(0\right)\left[1-G_{\psi}\left(0\right)\right]}=\frac{E^{\dagger}\int_{-\infty}^{+\infty}\left|\gamma\left(u;t\right)\right|^2 dF_{\psi}\left(u\right)}{G_{\psi}\left(0\right)\left[1-G_{\psi}\left(0\right)\right]}=$$
$$=E^{\dagger}\int_{-\infty}^{+\infty}\left|W^{-1/2}\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(u;t\right)-G_{\psi}\left(0\right)\right]\right|^2 dF_{\psi}\left(u\right)=$$
$$=W^{-1}\int_{-\infty}^{+\infty}E^{\dagger}\left|\widetilde{\zeta}_j\left(u;t\right)-G_{\psi}\left(0\right)\right|^2 dF_{\psi}\left(u\right);$$

by similar passages as in the proof of Theorem 1, it can be shown that

$$\frac{E^{\dagger}\left(\Gamma_t\right)}{G_{\psi}\left(0\right)\left[1-G_{\psi}\left(0\right)\right]}-1\leq C_0\left[Wh^{-2}\left(\frac{R}{\ln R}\right)+h^{-1}\left(\frac{R}{\ln R}\right)\right].$$

Thus

$$\max_{1\leq k\leq T_m}\sqrt{\frac{m}{k\left(k+m\right)}}\left|\sum_{t=m+1}^{m+k}\left(E^{\dagger}\left(\Gamma_t\right)-1\right)\right|\leq$$
$$\leq C_0W\left(\frac{\ln R}{R}\right)^2\max_{1\leq k\leq T_m}\sqrt{\frac{m}{k\left(k+m\right)}}k\ \leq$$
$$\leq C_0m^{1/2}\left[Wh^{-2}\left(\frac{R}{\ln R}\right)+h^{-1}\left(\frac{R}{\ln R}\right)\right],$$

which is $O\left(m^{-\epsilon}\right)$ on account of Assumption 9*(ii)*.

We now turn to (A.11). Let $\gamma\left(0;t\right)=W^{-1/2}\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(0;t\right)-G_{\psi}\left(0\right)\right]$; we have

$$\left(\int_{-\infty}^{+\infty}\left|\gamma\left(u;t\right)\right|^2 dF_{\psi}\left(u\right)\right)^2-\left(\int_{-\infty}^{+\infty}\left|\gamma\left(0;t\right)\right|^2 dF_{\psi}\left(u\right)\right)^2= \tag{A.13}$$
$$=\int_{-\infty}^{+\infty}\left(\left|W^{-1/2}\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(u;t\right)-\widetilde{\zeta}_j\left(0;t\right)\right]\right|^2+\right.$$
$$\left.+2W^{-1}\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(u;t\right)-\widetilde{\zeta}_j\left(0;t\right)\right]\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(0;t\right)-G_{\psi}\left(0\right)\right]^2\right)^2 dF_{\psi}\left(u\right)+$$
$$+\,2\int_{-\infty}^{+\infty}W^{-1/2}\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(0;t\right)-G_{\psi}\left(0\right)\right]\times\left(\left|W^{-1/2}\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(u;t\right)-\widetilde{\zeta}_j\left(0;t\right)\right]\right|^2+\right.$$
$$\left.+\,2W^{-1}\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(u;t\right)-\widetilde{\zeta}_j\left(0;t\right)\right]\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(0;t\right)-G_{\psi}\left(0\right)\right]\right)dF_{\psi}\left(u\right).$$

By Rosenthal's inequality

$$
\begin{aligned}
& W^{-2}E^{\dagger}\left|\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(u;t\right)-\widetilde{\zeta}_j\left(0;t\right)\right]\right|^4 \leq \qquad (A.14)\\
& \leq C_0W^{-2}\left[\left|\sum_{j=1}^{W}E^{\dagger}\left(\widetilde{\zeta}_j\left(u;t\right)-\widetilde{\zeta}_j\left(0;t\right)\right)\right|^4+\sum_{j=1}^{W}E^{\dagger}\left|\widetilde{\zeta}_j\left(u;t\right)-\widetilde{\zeta}_j\left(0;t\right)\right|^4\right] \leq \\
& \leq C_0W^{-2}\left[W^4\left|E^{\dagger}\left(\widetilde{\zeta}_j\left(u;t\right)-\widetilde{\zeta}_j\left(0;t\right)\right)\right|^4+C_1W\right] \leq \\
& \leq C_0W^{-2}\left[W^4\left|G_{\psi}\left(\psi_{N,m,R}^{-1}\left(t\right)\right)-G_{\psi}\left(0\right)\right|^4+C_1W\right] \leq \\
& \leq C_0W^{-1}+C_1W^2u^4\psi_{N,m,R}^{-4}\left(t\right);
\end{aligned}
$$

the same logic also yields

$$
\begin{aligned}
& W^{-1}E^{\dagger}\left|\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(u;t\right)-\widetilde{\zeta}_j\left(0;t\right)\right]\right|^2 \leq C_0u\psi_{N,m,R}^{-1}\left(t\right)+C_1Wu^2\psi_{N,m,R}^{-2}\left(t\right),\\
& E^{\dagger}\left|W^{-1/2}\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(0;t\right)-G_{\psi}\left(0\right)\right]\right|^2 \leq C_0.
\end{aligned}
$$

Repeated applications of the C_r-inequality and of the Cauchy-Schwartz inequality to (A.13) yield

$$
E^{\dagger}\left(\int_{-\infty}^{+\infty}\left|\gamma\left(u;t\right)\right|^2dF_{\psi}\left(u\right)\right)^2-E^{\dagger}\left(\int_{-\infty}^{+\infty}\left|\gamma\left(0;t\right)\right|^2dF_{\psi}\left(u\right)\right)^2 \leq C_0W\psi_{N,m,R}^{-2}\left(t\right).
$$

Finally, tedious but elementary calculations yield

$$
E^{\dagger}\left|W^{-1/2}\sum_{j=1}^{W}\left[\widetilde{\zeta}_j\left(0;t\right)-G_{\psi}\left(0\right)\right]\right|^4=3\left(G_{\psi}\left(0\right)\left[1-G_{\psi}\left(0\right)\right]\right)^2+O\left(W^{-1}\right).
$$

Putting all together and using (A.10), it follows that

$$
V^{\dagger}\left(\Gamma_t\right)-2 \leq C_0\left(W\psi_{N,m,R}^{-2}\left(t\right)+W^{-1}\right),
$$

whence (A.11) follows.

Consider now (A.12). By convexity

$$
\left|\int_{-\infty}^{+\infty}\left|\gamma\left(u;t\right)\right|^2dF_{\psi}\left(u\right)\right|^{2+\delta} \leq \int_{-\infty}^{+\infty}\left|\gamma\left(u;t\right)\right|^{4+2\delta}dF_{\psi}\left(u\right);
$$

also, applying the C_r-inequality

$$\int_{-\infty}^{+\infty} |\gamma(u;t)|^{4+2\delta} dF_\psi(u) \leq$$
$$\leq C_0 \int_{-\infty}^{+\infty} \left| W^{-1/2} \sum_{j=1}^{W} \left[\widetilde{\zeta}_j(u;t) - G_\psi\left(u\psi_{N,m,R}^{-1}(t)\right)\right] \right|^{4+2\delta} dF_\psi(u) +$$
$$+ C_0 \int_{-\infty}^{+\infty} \left| W^{-1/2} \sum_{j=1}^{W} \left[G_\psi\left(u\psi_{N,m,R}^{-1}(t)\right) - G_\psi(0)\right] \right|^{4+2\delta} dF_\psi(u).$$

Note that, by applying Burkholder's inequality and convexity

$$\int_{-\infty}^{+\infty} E^\dagger \left| W^{-1/2} \sum_{j=1}^{W} \left[\widetilde{\zeta}_j(u;t) - G_\psi\left(u\psi_{N,m,R}^{-1}(t)\right)\right] \right|^{4+2\delta} dF_\psi(u) \leq$$
$$\leq \int_{-\infty}^{+\infty} E^\dagger \left| W^{-1} \sum_{j=1}^{W} \left[\widetilde{\zeta}_j(u;t) - G_\psi\left(u\psi_{N,m,R}^{-1}(t)\right)\right]^2 \right|^{2+\delta} dF_\psi(u) \leq$$
$$\leq W^{-1} \sum_{j=1}^{W} \int_{-\infty}^{+\infty} E^\dagger \left| \left[\widetilde{\zeta}_j(u;t) - G_\psi\left(u\psi_{N,m,R}^{-1}(t)\right)\right]^2 \right|^{2+\delta} dF_\psi(u) \leq C_0,$$

since $\widetilde{\zeta}_j(u;t)$ has finite moments of any order. Also

$$\int_{-\infty}^{+\infty} \left| W^{-1/2} \sum_{j=1}^{W} \left[G_\psi\left(u\psi_{N,m,R}^{-1}(t)\right) - G_\psi(0)\right] \right|^{4+2\delta} dF_\psi(u) \leq$$
$$\leq W^{2+\delta} \int_{-\infty}^{+\infty} \left| G_\psi\left(u\psi_{N,m,R}^{-1}(t)\right) - G_\psi(0)\right|^{4+2\delta} dF_\psi(u) \leq$$
$$\leq W^{2+\delta} \left(\frac{m_{G,\psi}}{\psi_{N,m,R}(t)} \right)^{4+2\delta} \int_{-\infty}^{+\infty} |u|^{4+2\delta} dF_\psi(u) \leq C_0,$$

where $m_{G,\psi}$ is the upper bound of the density of $G_\psi(\cdot)$; the final estimate follows from Assumptions 7 and 9*(i)*. This proves (A.12). □

Appendix A.2. Proofs of main results

Proof of Lemma 1. Recall the definition

$$\Sigma_m(t) = \frac{1}{m} \sum_{k=t-m+1}^{t} E\left(X_k X_k'\right).$$

When $m \leq t < \tau$ and under $H_{A,1}$ or $H_{A,2}$, it is easy to see that there is no change in the $(r+1)$-th eigenvalue of $E(X_k X_k')$. Thus, the proof that $\lambda^{(r+1)}(t)$ is finite is exactly the same as the proof of Lemma 2.1 in Trapani [62]. The same holds under H_0 for all $m \leq t \leq T$.

We begin with studying $H_{A,2}$. When $\tau \leq t < \tau + m - 1$, it holds that

$$\begin{aligned}\Sigma_m(t) &= \frac{1}{m}\sum_{k=t-m+1}^{\tau-1} E\left(X_k X_k'\right) + \frac{1}{m}\sum_{k=\tau}^{t} E\left(X_k X_k'\right) \\ &= \frac{\tau+m-t-1}{m}\Sigma_m^{(1)}(t) + \frac{t-\tau+1}{m}\Sigma_m^{(2)}(t). \end{aligned} \tag{A.15}$$

Let $\lambda_1^{(r+1)}(t)$ and $\lambda_2^{(r+1)}(t)$ be the $(r+1)$-th eigenvalue of $\Sigma_m^{(1)}(t)$ and $\Sigma_m^{(2)}(t)$ respectively. Now $\lambda_1^{(r+1)}(t)$ depends only on observations before the break, therefore $\lambda_1^{(r+1)}(t) < \infty$: this can be shown again by following the proof of Lemma 2.1 in Trapani [62]. As far as $\lambda_2^{(r+1)}(t)$ is concerned, it depends only on post-break observations, which are driven by a factors vector f_t of size $(r+q)$. In particular, $\Sigma_m^{(2)}(t) = A(t)E\left(f_t f_t'\right)A(t)' + \Sigma_u(t)$, where $A(t) = [A|B]$ is a constant $N \times (r+q)$ matrix. By Weyl's inequality and Assumption 2

$$\lambda_2^{(r+1)}(t) \geq \gamma^{(r+1)}(t) + \omega^{(N)}(t) \geq \underline{C}_{r+1}(t)N. \tag{A.16}$$

Then, applying Weyl's inequality and (A.16) to (A.15), it follows that

$$\lambda^{(r+1)}(t) \geq \frac{\tau+m-t-1}{m}\lambda_1^{(\min)}(t) + \frac{t-\tau+1}{m}\lambda_2^{(r+1)}(t),$$

which yields (10) immediately. When instead $\tau + m - 1 \leq t \leq T$ we have $\Sigma_m(t) = \Sigma_m^{(2)}(t)$ and the result follows directly from (A.16). Under $H_{A,1}$, $r(t) = r$ for $m \leq t \leq T$ and therefore $\Sigma_F(t) = E(f_t f_t')$ is a $r \times r$ constant matrix and we denote it as Σ_F. We have

$$\Sigma_X(t) = \left(A\Sigma_F A'\right) I_{[m,\tau)}(t) + \left(\widetilde{A}\Sigma_F \widetilde{A}'\right) I_{[\tau,T]}(t) + \Sigma_u(t). \tag{A.17}$$

Therefore, it holds that for $\tau \leq t < \tau + m - 1$

$$\Sigma_m(t) = \frac{1}{m}\sum_{k=t-m+1}^{t} \Sigma_X(k) = A^*\Sigma_F^*(t)A^{*\prime} + \Sigma_u(t),$$

where we have defined $A^* = [A|\widetilde{A}]$ and

$$\Sigma_F^*(t) = \begin{bmatrix} \frac{\tau+m-t-1}{m}\Sigma_F & 0 \\ 0 & \frac{t-\tau+1}{m}\Sigma_F \end{bmatrix}, \tag{A.18}$$

with the off-diagonal blocks being $r \times r$ matrices of zeros. Denoting by $\nu^{(k)}(\cdot)$ the k-th largest eigenvalues of a matrix, by Weyl's inequality, we have

$$\lambda^{(r+1)}(t) \geq \nu^{(r+1)}\left(A^*\Sigma_F^*(t)A^{*\prime}\right) + \omega^{(N)}(t) \geq \nu^{(r+1)}\left(A^*\Sigma_F^*(t)A^{*\prime}\right), \tag{A.19}$$

Now, since the spectrum of a block diagonal matrix is the union of the spectra of the blocks

$$\begin{aligned}
\nu^{(r+1)}\left(A^{*}\Sigma_F^{*}(t)A^{*\prime}\right) &= \nu^{(r+1)}\begin{pmatrix} \frac{\tau+m-t-1}{m}A\Sigma_F A' & 0 \\ 0 & \frac{t-\tau+1}{m}\widetilde{A}\Sigma_F\widetilde{A}' \end{pmatrix} \\
&= \min\left\{\nu^{(r)}\left(\frac{\tau+m-t-1}{m}A\Sigma_F A'\right), \nu^{(r)}\left(\frac{t-\tau+1}{m}\widetilde{A}\Sigma_F\widetilde{A}'\right)\right\} \\
&\geq \min\left\{\frac{\tau+m-t-1}{m}, \frac{t-\tau+1}{m}\right\} \min\left\{\nu^{(r)}\big(A\Sigma_F A'\big), \nu^{(r)}\big(\widetilde{A}\Sigma_F\widetilde{A}'\big)\right\} \\
&\geq \min\left\{\frac{\tau+m-t-1}{m}, \frac{t-\tau+1}{m}\right\} \underline{C}_r N, \qquad \text{(A.20)}
\end{aligned}$$

where the last inequality follows from the fact that by (A.17) and Assumption 2 we have

$$\gamma^{(r)}(t) = \nu^{(r)}\big(A\Sigma_F A'\big)I_{[m,\tau)}(t) + \nu^{(r)}\big(\widetilde{A}\Sigma_F\widetilde{A}'\big)I_{[\tau,T]}(t) \geq \underline{C}_r N,$$

which implies that $\nu^{(r)}\big(A\Sigma_F A'\big) \geq \underline{C}_r N$ and $\nu^{(r)}\big(\widetilde{A}\Sigma_F\widetilde{A}'\big) \geq \underline{C}_r N$. Using (A.20) in (A.19), equation (9) follows. Last, when $\tau+m-1 \leq t \leq T$, we have $A^{*} = \widetilde{A}$ and $\Sigma_F^{*}(t) = \Sigma_F$ and the proof is the same as in Lemma 2.1 in Trapani [62]. □

Proof of Lemma 2. The proof is essentially the same as the proof of Lemma 2.2 in Trapani [62], and we report it here in full for completeness. Consider the eigenvalue stability inequality (see e.g. Horn and Johnson [35], p. 367), viz.

$$\left|\widehat{\lambda}^{(r+1)}(t) - \lambda^{(r+1)}(t)\right| \leq \left\|\widehat{\Sigma}_m(t) - \Sigma_m(t)\right\|_{op}, \qquad \text{(A.21)}$$

where $\|\cdot\|_{op}$ is the operator norm. By symmetry, $\left\|\widehat{\Sigma}_m(t) - \Sigma_m(t)\right\|_{op} \leq \left\|\widehat{\Sigma}_m(t) - \Sigma_m(t)\right\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm; hence, (A.21) becomes

$$\left|\widehat{\lambda}^{(r+1)}(t) - \lambda^{(r+1)}(t)\right| \leq \left[\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{1}{m}\sum_{k=t-m+1}^{t} X_{i,k}X_{j,k} - E(X_{i,k}X_{j,k})\right)^2\right]^{1/2}. \qquad \text{(A.22)}$$

We now provide an estimate for $\left|\widehat{\lambda}^{(r+1)}(t) - \lambda^{(r+1)}(t)\right|$; the proof uses, in a multi-index context, the same approach as Cai [18]. Let $\delta_{h,j,k} \equiv X_{h,k}X_{j,k} - E(X_{h,k}X_{j,k})$, and $t_0 = t-m+1$; we begin by showing

$$\begin{aligned}
\sum_{N=1}^{\infty}\sum_{N=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{N^2 m}P\Bigg[\max_{1\leq\tilde{h}\leq N, 1\leq\tilde{j}\leq N, t_0\leq\tilde{t}\leq t_0+m-1}\left|\sum_{h=1}^{\tilde{h}}\sum_{j=1}^{\tilde{j}}\left(\frac{1}{m}\sum_{k=t_0}^{\tilde{t}}\delta_{h,j,k}\right)^2\right|^{1/2} \\
> \varepsilon\frac{N}{\sqrt{m}}\ln^{1+\epsilon}N\ln^{\frac{1+\epsilon}{2}}m\Bigg] < \infty, \qquad \text{(A.23)}
\end{aligned}$$

for some $\varepsilon > 0$ and any $\epsilon > 0$. Equation (A.23) can be shown by noting that

$$E\left[\max_{1\leq\tilde{h}\leq N,1\leq\tilde{j}\leq N,t_0\leq\tilde{t}\leq t_0+m-1}\left|\sum_{h=1}^{\tilde{h}}\sum_{j=1}^{\tilde{j}}\left(\frac{1}{m}\sum_{k=t_0}^{\tilde{t}}\delta_{h,j,k}\right)^2\right|\right]$$
$$\leq\sum_{h=1}^{N}\sum_{j=1}^{N}E\left[\max_{t_0\leq\tilde{t}\leq t_0+m-1}\left|\left(\frac{1}{m}\sum_{k=t_0}^{\tilde{t}}\delta_{h,j,k}\right)^2\right|\right]\leq C_0\frac{N^2}{m},$$

by virtue of Assumption 3*(ii)*. Thus, by Markov inequality, (A.23) holds since

$$\sum_{N=1}^{\infty}\sum_{N=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{N^2m}\frac{m}{\varepsilon^2N^2\left[\ln^{1+\epsilon}N\ln^{\frac{1+\epsilon}{2}}m\right]^2}C_0\frac{N^2}{m}<\infty.$$

We now show that (A.23) entails

$$\lim\sup_{N,m\to\infty}\frac{\left|\sum_{h=1}^{N}\sum_{j=1}^{N}\left(\sum_{k=t_0}^{t_0+m-1}\delta_{h,j,k}\right)^2\right|^{1/2}}{N\sqrt{m}\ln^{1+\epsilon}N\ln^{\frac{1+\epsilon}{2}}m}=0\text{ a.s.}\tag{A.24}$$

Similarly to the proof of Lemma A.2, note that for every triple (N,N,m), there is a triple of positive integers (k_1',k_2',k_3') such that $2^{k_1'}\leq N<2^{k_1'+1}$, $2^{k_2'}\leq N<2^{k_2'+1}$, $2^{k_3'}\leq m<2^{k_3'+1}$. Further, there is also a triple of real numbers defined over $[0,1)$, say $(\rho_1',\rho_2',\rho_3')$, such that $N=2^{k_1'+\rho_1'}$, etc... Define now the short-hand notation

$$L\left(k_1',k_2',k_3'\right)\equiv\frac{\sqrt{2^{k_1'+1}}\sqrt{2^{k_2'+1}}}{\sqrt{2^{k_3'+1}}}\ln^{\frac{1+\epsilon}{2}}\left(2^{k_1'+\rho_1'}\right)\ln^{\frac{1+\epsilon}{2}}\left(2^{k_2'+\rho_2'}\right)\ln^{\frac{1+\epsilon}{2}}\left(2^{k_3'+\rho_3'}\right),$$
$$S\left(k_1',k_2',k_3'\right)\equiv\left|\sum_{h=1}^{k_1'}\sum_{j=1}^{k_2'}\left(\sum_{k=t_0}^{t_0+k_3'-1}\delta_{h,j,k}\right)^2\right|^{1/2},$$
$$P_{k_1',k_2',k_3'}\equiv P\left[\max_{1\leq k_1'\leq 2^{k_1'+\rho_1'},1\leq k_2'\leq 2^{k_2'+\rho_2'},1\leq k_3'\leq 2^{k_3'+\rho_3'}}\left|S\left(k_1',k_2',k_3'\right)\right|>\varepsilon L\left(k_1',k_2',k_3'\right)\right].$$

Equation (A.23) entails that

$$\sum_{k_1'=0}^{\infty}\sum_{k_2'=0}^{\infty}\sum_{k_3'=0}^{\infty}\frac{2^{k_1'}2^{k_2'}2^{k_3'}}{\left(2^{k_1'+1}-1\right)\left(2^{k_2'+1}-1\right)\left(2^{k_3'+1}-1\right)}P_{k_1',k_2',k_3'}<\infty;$$

thus

$$\sum_{k_1'=0}^{\infty}\sum_{k_2'=0}^{\infty}\sum_{k_3'=0}^{\infty}P_{k_1',k_2',k_3'}\leq 2^3\sum_{k_1'=0}^{\infty}\sum_{k_2'=0}^{\infty}\sum_{k_3'=0}^{\infty}\frac{2^{k_1'}2^{k_2'}2^{k_3'}}{\left(2^{k_1'+1}-1\right)\left(2^{k_2'+1}-1\right)\left(2^{k_3'+1}-1\right)}P_{k_1',k_2',k_3'}<\infty,$$

so that the Borel-Cantelli Lemma yields

$$\frac{\max_{k_1',k_2',k_3'}\left|S\left(k_1',k_2',k_3'\right)\right|}{L\left(k_1',k_2',k_3'\right)}\to 0 \text{ a.s.},$$

whence

$$\frac{\left|S\left(N,N,T\right)\right|}{L\left(N,N,T\right)}\leq\frac{\max_{k_1',k_2',k_3'}\left|S\left(k_1',k_2',k_3'\right)\right|}{L\left(k_1',k_2',k_3'\right)}\frac{L\left(k_1',k_2',k_3'\right)}{L\left(N,N,T\right)}\leq\sqrt{2}\frac{\max_{k_1',k_2',k_3'}\left|S\left(k_1',k_2',k_3'\right)\right|}{L\left(k_1',k_2',k_3'\right)}\to 0 \text{ a.s.},$$

so that finally

$$\lim\sup_{N,T\to\infty}\frac{\left|S\left(N,N,T\right)\right|}{L\left(N,N,T\right)}=0 \text{ a.s.}.$$

The desired result now follows immediately.

□

Proof of Theorem 1. Consider (15); we report its proof, which is a refinement of the proof of Theorem 3 in Horváth and Trapani [36], in full. In the presence of a break, it follows from Lemmas 1 and 2 that

$$P\left\{\omega:\lim_{N,m\to\infty}\phi_{N,m}\left(t\right)=\infty\right\}=1,$$

for each $t\geq\tau$, as long as

$$\begin{aligned} N^{1-\delta}\min\left\{\frac{t-\tau+1}{m},\frac{\tau+m-t-1}{m}\right\} &\to \infty \text{ for } \tau\leq t<\tau+m-1,\\ N^{1-\delta} &\to \infty \text{ for } t\geq\tau+m-1,\end{aligned}$$

hold. By definition, this holds true within the intervals $t_{N,m}^{*}\leq t\leq t_{N,m}^{*}$ under $H_{A,1}$, and $t_{N,m}^{*}\leq t\leq T$ under $H_{A,2}$. Thus, we can assume from now on that $\lim_{N,m\to\infty}\phi_{N,m}\left(t\right)=\infty$ holds in the prescribed intervals. Note that

$$\begin{aligned} R^{-1/2}\sum_{i=1}^{R}\left[\zeta_i(u;t)-G_\phi\left(0\right)\right] &= R^{-1/2}\sum_{i=1}^{R}\left[I\{\xi_i\leq 0\}-G_\phi\left(0\right)\right]+\\ &+R^{-1/2}\sum_{i=1}^{R}\left[I\{\xi_i\leq u\phi_{N,m}^{-1}\left(t\right)\}-I\{\xi_i\leq 0\}-(G_\phi(u\phi_{N,m}^{-1}\left(t\right))-G_\phi(0))\right]+\\ &+R^{-1/2}\sum_{i=1}^{R}\left[G_\phi(u\phi_{N,m}^{-1}\left(t\right))-G_\phi(0)\right]. \end{aligned} \tag{A.25}$$

By construction,

$$\begin{aligned} E^{*}\zeta_j(u;t) &= G_\phi(u\phi_{N,m}^{-1}\left(t\right)),\\ E^{*}(\zeta_j(u;t)-E^{*}\zeta_j(u;t))^2 &= G_\phi(u\phi_{N,m}^{-1}\left(t\right))\left[1-G_\phi(u\phi_{N,m}^{-1}\left(t\right))\right].\end{aligned}$$

Consider now the following passages:

$$E^*\left[\int_{-\infty}^{\infty}\left(R^{-1/2}\sum_{i=1}^{R}\left[I\{\xi_i \le u\phi_{N,m}^{-1}(t)\} - I\{\xi_i \le 0\} - (G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0))\right]\right)^2 dF_\phi(u)\right] =$$
$$= \int_{-\infty}^{\infty} E^*\left[I\{\xi_1 \le u\phi_{N,m}^{-1}(t)\} - I\{\xi_1 \le 0\} - (G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0))\right]^2 dF_\phi(u),$$

on account of the independence of the ξ_is. Also, note that the random variable $I\{\xi_1 \le u\phi_{N,m}^{-1}(t)\} - I\{\xi_1 \le 0\}$ has expected value given by $G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0)$, and variance equal to

$$\begin{aligned}
&E^*\left[I\{\xi_1 \le u\phi_{N,m}^{-1}(t)\} - I\{\xi_1 \le 0\} - G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0)\right]^2 = \\
&= \left(G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0)\right)\left[1 - G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0)\right] \le \\
&\le G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0).
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} E^*\left[I\{\xi_1 \le u\phi_{N,m}^{-1}(t)\} - I\{\xi_1 \le 0\} - (G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0))\right]^2 dF_\phi(u) \le \\
&\le \int_{-\infty}^{\infty}\left[G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0)\right] dF_\phi(u) \le \frac{m_G}{\phi_{N,m}(t)}\int_{-\infty}^{\infty}|u|dF_\phi(u), \qquad \text{(A.26)}
\end{aligned}$$

where the last passage follows from Assumption 6*(i)*, with m_G an upper bound for the density of G. Also

$$\int_{-\infty}^{\infty}\left(R^{1/2}\left[G_\phi(u\phi_{N,m}^{-1}(t)) - G_\phi(0)\right]\right)^2 dF_\phi(u) \le \frac{R}{\phi_{N,m}^2(t)} m_G \int_{-\infty}^{\infty} u^2 dF_\phi(u). \qquad \text{(A.27)}$$

Hence, using Assumptions 4 and 5, we conclude via Markov's inequality that

$$\begin{aligned}
\Theta_t &= \int_{-\infty}^{\infty}\left\{\frac{1}{\sqrt{G_\phi(0)\left[1 - G_\phi(0)\right]}R^{1/2}}\sum_{i=1}^{R}\left[I\{\xi_i \le 0\} - G_\phi(0)\right]\right\}^2 dF_\phi(u) + o_{P^*}(1) = \\
&= \left\{\frac{1}{\sqrt{G_\phi(0)\left[1 - G_\phi(0)\right]}R^{1/2}}\sum_{i=1}^{R}\left[I\{\xi_i \le 0\} - G_\phi(0)\right]\right\}^2 + o_{P^*}(1),
\end{aligned}$$

and therefore the result follows from the Central Limit Theorem for Bernoulli random variables.

Equation (16) can be shown by exactly the same logic as the proof of (A.2), and of Theorem 4 in Horváth and Trapani [36], and is therefore omitted. □

Proof of Theorem 2. The proof of the theorem is exactly the same as that of Theorem 1, but this time based on Lemma A.2. □

Proof of Theorem 3. By (A.10) and (A.11), we have

$$\max_{1\leq k\leq T_m}\sqrt{\frac{m}{k\left(k+m\right)}}\left|\sum_{t=m+1}^{m+k}\frac{\Gamma_t-1}{\sqrt{2}}\right|=\max_{1\leq k\leq T_m}\sqrt{\frac{m}{k\left(k+m\right)}}\left|\sum_{t=m+1}^{m+k}Z_t\right|+O\left(m^{-\epsilon}\right),\quad\text{(A.28)}$$

where

$$Z_t=\frac{\Gamma_t-E^{\dagger}\left(\Gamma_t\right)}{\sqrt{V^{\dagger}\left(\Gamma_t\right)}}$$

is an *i.i.d.* sequence with mean zero, unit variance and finite moments of order $2+\delta$. Consider (26); on account of (A.28), this holds immediately, following the same passages as in the proof of Theorem 2.1 in Horváth *et al.* [37]. As far as (27) is concerned, due to the polynomial rate of approximation in (A.28), it suffices to prove that

$$P^{\dagger}\left(A_m\max_{1\leq k\leq T_m}\sqrt{\frac{m}{k\left(k+m\right)}}\left|\sum_{t=m+1}^{m+k}Z_t\right|\geq x+D_m\right)=e^{-e^{-x}},$$

as $\min\left(N,m,R,W\right)\to\infty$. This is a relatively standard exercise, and it is very similar to the proof of Theorem 1.1 in Horváth *et al.* [38]; we therefore report only the main passages. Let $a\left(m\right)=\left(\ln m\right)^2$; by virtue of (A.12), it holds that

$$\sup_{a(m)\leq k<\infty}\sqrt{\frac{m}{k\left(k+m\right)}}\left|\sum_{t=m+1}^{m+k}Z_t-B\left(k\right)\right|=O_{P^{\dagger}}\left(a\left(m\right)^{-\frac{\epsilon}{2(2+\epsilon)}}\right),\quad\text{(A.29)}$$

where $\left\{B\left(t\right),\ 0\leq t<\infty\right\}$ is a standard Wiener process – see Komlós *et al.* [43, 44]. We now show that

$$\max_{a(m)\leq k\leq\frac{cm}{\ln m}}\frac{B\left(\frac{k}{k+m}\right)}{\sqrt{\frac{k}{m}}}=\max_{a(m)\leq k\leq\frac{cm}{\ln m}}\frac{B\left(\frac{k}{k+m}\right)}{\sqrt{\frac{k}{k+m}}}+O_{P^{\dagger}}\left(\frac{\left(\ln\ln m\right)^{1/2}}{\ln m}\right);\quad\text{(A.30)}$$

given that

$$\left|\frac{k}{m}-\frac{k}{k+m}\right|\leq\left(\frac{k}{m}\right)^2,$$

using the modulus of continuity of the Wiener process we obtain

$$\max_{a(m)\leq k\leq\frac{cm}{\ln m}}\left|\frac{B\left(\frac{k}{k+m}\right)}{\sqrt{\frac{k}{m}}}-\frac{B\left(\frac{k}{k+m}\right)}{\sqrt{\frac{k}{k+m}}}\right|=O_{P^{\dagger}}\left(1\right)\max_{a(m)\leq k\leq\frac{cm}{\ln m}}\frac{k}{m}\left(\ln\frac{m}{k}\right)^{1/2},$$

whence (A.30) follows. Consequently, the following results hold:

$$\frac{1}{\sqrt{2\ln\ln m}}\max_{1\leq k\leq T_m}\frac{B\left(\frac{k}{k+m}\right)}{\sqrt{\frac{k}{m}}}\overset{P^\dagger}{\to}1, \tag{A.31}$$

$$A_m\max_{1\leq k\leq a(m)}\frac{B\left(\frac{k}{k+m}\right)}{\sqrt{\frac{k}{m}}}-D_m\overset{P^\dagger}{\to}-\infty, \tag{A.32}$$

$$A_m\max_{\frac{cm}{\ln m}\leq k\leq T_m}\frac{B\left(\frac{k}{k+m}\right)}{\sqrt{\frac{k}{m}}}-D_m\overset{P^\dagger}{\to}-\infty; \tag{A.33}$$

the results above are shown, for $\sqrt{\frac{k+m}{m}}B\left(\frac{k}{k+m}\right)$, in Lemmas 3.4, 3.5 and (in the proof of) Lemma 3.6 in Horváth *et al.* [38]; in (A.33) we have used the fact that, by Assumption 8, there exists a $c>0$ such that $T_m>cm$. Combining (A.29), (A.31), (A.32), (A.33) and (A.30) together, we obtain

$$P^\dagger\left(A_m\max_{1\leq k\leq T_m}\frac{1}{\sqrt{k\left(k+m\right)}}\left|\sum_{t=m+1}^{m+k}Z_t\right|\geq x+D_m\right)=$$
$$=P^\dagger\left(A_m\max_{a(m)\leq k\leq\frac{cm}{\ln m}}\frac{\left|B\left(\frac{k}{k+m}\right)\right|}{\sqrt{\frac{k}{k+m}}}\geq x+D_m\right)+o\left(1\right);$$

then the desired result follows from Lemma 3.6 in Horváth *et al.* [38].

Consider now (28); it is convenient to prove the result under $H_{A,2}$ first. On account of (9), Lemmas 1 and 2, Assumption 4*(i)* and (19), it holds that

$$\Gamma_t=C_0W+o_{P^\dagger}\left(W\right),\ \text{for } t\geq\tau+C_1m^{1/2+\epsilon},$$

where $\epsilon>0$ is such that $\frac{N^{1-\delta}}{m^{1/2-\epsilon}}\to C_2\in(0,+\infty)$ and

$$C_0=\int_{-\infty}^{+\infty}\frac{\left|G_\psi\left(u\right)-G_\psi\left(0\right)\right|^2}{G_\psi\left(0\right)\left[1-G_\psi\left(0\right)\right]}dF_\psi\left(u\right).$$

Thus, standard algebra yields that under $H_{A,2}$

$$\sum_{m+1}^{m+k}\frac{\Gamma_t-1}{\sqrt{2}}=O_{P^\dagger}\left(1\right)\left[m+k-\left(\tau+C_1m^{1/2+\epsilon}\right)\right]W+o_{P^\dagger}\left(W\right),$$

whenever $k\geq\tau+C_1m^{1/2+\epsilon}$. Therefore,

$$\Lambda_m=O_{P^\dagger}\left(1\right)m^{1/2}W\max_{1\leq k\leq T_m}\frac{m+k-\left(\tau+C_1m^{1/2+\epsilon}\right)}{k^{1/2}\left(m+k\right)^{1/2}}+o_{P^\dagger}\left(W\right);$$

elementary algebra yields

$$\max_{1\leq k\leq T_m}\frac{m+k-\left(\tau+C_1m^{1/2+\epsilon}\right)}{k^{1/2}\left(m+k\right)^{1/2}}\geq C_2>0;$$

thus, Assumption 8 implies (28). Under $H_{A,1}$ the logic is similar, and therefore only the main passages are reported. Under $H_{A,1}$ we have

$$\Gamma_t=C_0W+o_{P^\dagger}\left(W\right),\ \text{for}\ \tau+C_1m^{1/2+\epsilon}\leq t\leq\tau+m-C_1m^{1/2+\epsilon},$$

with the same notation as above; it is then easy to see that

$$\max_{1\leq k\leq T_m}\frac{m+k-\left(\tau+C_1m^{1/2+\epsilon}\right)}{k^{1/2}\left(m+k\right)^{1/2}}\geq C_2>0;$$

the proof is now the same as before. □

Proof of Corollary 1. The corollary is an immediate consequence of Theorem 3 and its proof. Considering (29), note that $P^\dagger\left(\widehat{\tau}_m<T\right)$ is monotonically nondecreasing in T; by definition

$$\begin{aligned}P^\dagger\left(\widehat{\tau}_m<T\right)&=P\left(\max_{1\leq k\leq T}\frac{d\left(k;m\right)}{\nu^*\left(k;m\right)}>c_{\alpha,m}\right)\leq P\left(\max_{1\leq k<\infty}\frac{d\left(k;m\right)}{\nu^*\left(k;m\right)}>c_{\alpha,m}\right)=\\&=P\left(\sup_{0\leq t\leq 1}\frac{\left|B\left(t\right)\right|}{t^\eta}>c_{\alpha,m}\right)+o\left(1\right)=\alpha,\end{aligned}$$

which proves (26); (27) follows from the same passages. Similarly, as far as (30) is concerned, note that

$$P^\dagger\left(t^*_{N,m}\leq\widehat{\tau}_m<T\right)=P\left(c^{-1}_{\alpha,m}\max_{1\leq k\leq T_m}\frac{d\left(k;m\right)}{\nu^*\left(k;m\right)}>1\right)=1,$$

by (28). □