Submitted to the Annals of Statistics

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

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

Shandong University [‡], *Nanjing Audit University* [§], *University of Leicester* [¶], *and National University of Singapore*[‖]

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

[†]Corresponding author.

CONTENTS

A. Assumptions and discussion. The following assumptions are borrowed (and, in some cases mentioned below, adapted) from the papers by Yu et al. (2022) and He et al. (2023), to which we refer for (further) detailed explanations and discussions. Henceforth, we will make use of the following notation, which complements the one already presented in the main paper: given a matrix $\mathbf{A}$, $\|\mathbf{A}\|_{\max}$ represents $\max_{1\le i\le n_1, 1\le j\le n_2}|A_{ij}|$; and, with obvious notation, $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote, respectively, the smallest and largest eigenvalues of $\mathbf{A}$. Finally, we use "$\otimes$" to denote the Kronecker product.

ASSUMPTION A.1. *For some $r \ge 2$, it holds that: (i) (a) $E(\mathbf{F}_t) = \mathbf{0}_{k_1\times k_2}$, and (b) $E\|\mathbf{F}_t\|^{4r} \le c_0$; (ii)*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t' \xrightarrow{a.s.} \mathbf{\Sigma}_1 \text{ and } \frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t'\mathbf{F}_t \xrightarrow{a.s.} \mathbf{\Sigma}_2, \tag{A.1}$$

where Σ_i is a $k_i \times k_i$ positive definite matrix, $i = 1, 2$. The factor numbers k_1 and k_2 are fixed as $\min\{p_1, p_2, T\} \to \infty$; (iii) it holds that, for all $1 \le h_1, l_1 \le k_1$ and $1 \le h_2, l_2 \le k_2$

$$E\max_{1\le\widetilde{t}\le T}\left(\sum_{t=1}^{\widetilde{t}}\left(F_{h_1h_2,t}F_{l_1l_2,t} - E\left(F_{h_1h_2,t}F_{l_1l_2,t}\right)\right)\right)^{r} \le c_0T^{r/2}.$$

ASSUMPTION A.2. *(i) $\|\mathbf{R}\|_{\max} \le c_0$, and $\|\mathbf{C}\|_{\max} \le c_1$; (ii) as $\min\{p_1, p_2\} \to \infty$, $\|p_1^{-1}\mathbf{R}'\mathbf{R} - I_{k_1}\| \to 0$ and $\|p_2^{-1}\mathbf{C}'\mathbf{C} - I_{k_2}\| \to 0$.*

Assumptions A.1 and A.2 are standard in large factor models, and we refer, for example, to Chen and Fan (2021). In Assumption A.1*(i)*(b), however, we require the existence of (at least) the 8-th moment of $\mathbf{F}_t$, as opposed to the standard 4-th moment requirement (Bai, 2003). This is required in order to implement the projected estimator of Yu et al. (2022) (note that, in that paper, the restriction $E\|\mathbf{F}_t\|^4 \le c_0$ is explicitly required but, upon inspecting other assumptions, this needs to be strengthened to $E\|\mathbf{F}_t\|^8 \le c_0$ at least). Compared to Yu et al. (2022), Assumption A.1 is actually milder, since it does not require e.g. the eigenvalues of Σ_i to be distinct; indeed, this would be required in order to derive the limiting distribution of the estimated $\mathbf{R}$ and $\mathbf{C}$, but in our case we do not need any limiting distribution, and we only rely on rates. Part *(iii)* of the assumption is a maximal inequality, and it is the same as in He et al. (2023). This part of the assumption could be derived from more primitive dependence assumptions: for example, it can be shown to hold under various mixing conditions (see e.g. Rio, 1995; and Shao, 1995), and for the very general class of decomposable Bernoulli shifts (see e.g. Berkes et al., 2011; Liu and Lin, 2009; and Barigozzi and Trapani, 2022).

Assumption A.2 states that the common factors are pervasive (Uematsu and Yamagata, 2021). Whilst the main results in our paper are derived under this assumption, in Section B.1.3 we consider extension to the case of weak factors, viz. to the case $\|\mathbf{R}'\mathbf{R}\| = O\left(p_1^{\alpha}\right)$, with $\alpha < 1$.

ASSUMPTION A.3. *For some $r \ge 2$, it holds that: (i) (a) $E(e_{ij,t}) = 0$, and (b) $E\left|e_{ij,t}\right|^{4r} \le c_0$; (ii) for all $1 \le t \le T$, $1 \le i \le p_1$ and $1 \le j \le p_2$,*

$$(a).\sum_{s=1}^{T}\sum_{l=1}^{p_1}\sum_{h=1}^{p_2}\left|E(e_{ij,t}e_{lh,s})\right| \le c_0, \quad (b).\sum_{l=1}^{p_1}\sum_{h=1}^{p_2}\left|E(e_{lj,t}e_{ih,t})\right| \le c_0;$$

(iii) it holds that

$$
\begin{aligned}
&\text{(a) } E\left|\sum_{t=1}^{T}\left(e_{ij,t}e_{hk,t}-E\left(e_{ij,t}e_{hk,t}\right)\right)\right|^{2r} &\leq\ & c_0 T^r \text{ for all } 1\leq i,h\leq p_1 \text{ and } 1\leq j,k\leq p_2,\\
&\text{(b) } E\left|\sum_{s=1}^{T}\sum_{j=1}^{p_2}\left(e_{ij,t}e_{hk,t}-E\left(e_{ij,t}e_{hk,t}\right)\right)\right|^{2r} &\leq\ & c_0\left(Tp_2\right)^r \text{ for all } 1\leq i,h\leq p_1 \text{ and } 1\leq k\leq p_2,\\
&\text{(c) } E\left|\sum_{s=1}^{T}\sum_{i=1}^{p_1}\left(e_{ij,t}e_{hk,t}-E\left(e_{ij,t}e_{hk,t}\right)\right)\right|^{2r} &\leq\ & c_0\left(Tp_1\right)^r \text{ for all } 1\leq h\leq p_1 \text{ and } 1\leq j,k\leq p_2,\\
&\text{(d) } E\left|\sum_{s=1}^{T}\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\left(e_{ij,t}e_{hk,t}-E\left(e_{ij,t}e_{hk,t}\right)\right)\right|^{r} &\leq\ & c_0\left(Tp_1p_2\right)^{r/2} \text{ for all } 1\leq h\leq p_1 \text{ and } 1\leq k\leq p_2,\\
&\text{(d') } E\left|\sum_{s=1}^{T}\sum_{i=1}^{p_1}\sum_{k=1}^{p_2}\left(e_{ij,t}e_{hk,t}-E\left(e_{ij,t}e_{hk,t}\right)\right)\right|^{r} &\leq\ & c_0\left(Tp_1p_2\right)^{r/2} \text{ for all } 1\leq h\leq p_1 \text{ and } 1\leq j\leq p_2;
\end{aligned}
$$

(iv) it holds that $\lambda_{\min}\left[E\left(\frac{1}{p_2T}\sum_{t=1}^{T}\mathbf{E}_t\mathbf{E}_t'\right)\right]>0$ *and* $\lambda_{\min}\left[E\left(\frac{1}{p_1T}\sum_{t=1}^{T}\mathbf{E}_t'\mathbf{E}_t\right)\right]>0$.

Assumption A.3 ensures, in essence, the (cross-sectional and time series) summability of the idiosyncratic terms $e_{ij,t}$. Part *(iii)* of the assumption, in particular, controls the growth rate of the partial sums. As mentioned above for Assumption A.1, the growth rates for sums across time can be derived from more primitive assumptions; indeed, our assumption can be read in conjunction with the paper by Wang et al. (2019), where $\mathbf{E}_t$ is assumed to be white noise, which can be viewed as overly restrictive. Similar considerations would hold for summations across the cross-sectional dimensions; in this case, finding an ordering is less straightforward but, in principle, the assumption can be shown to hold under weak cross-sectional dependence. As a final remark, we note that part *(iii)(d)-(d')* of our assumption slightly modifies the assumptions in Yu et al. (2022) and He et al. (2023), where it is required that

$$
\begin{aligned}
\sum_{s=1}^{T}\sum_{l_2=1}^{p_1}\sum_{h=1}^{p_2}\left|Cov(e_{ij,t}e_{l_1j,t},e_{ih,s}e_{l_2h,s})\right| &\leq c_0,\\
\sum_{s=1}^{T}\sum_{l=1}^{p_1}\sum_{h_2=1}^{p_2}\left|Cov(e_{ij,t}e_{ih_1,t},e_{lj,s}e_{lh_2,s})\right| &\leq c_0,\\
\sum_{s=1}^{T}\sum_{l=1}^{p_1}\sum_{h=1}^{p_2}\left|Cov(e_{ij,t}^2,e_{lh,s}^2)\right| &\leq c_0,\\
\sum_{s=1}^{T}\sum_{l_2=1}^{p_1}\sum_{h_2=1}^{p_2}\left|Cov(e_{ij,t}e_{l_1h_1,t},e_{ij,s}e_{l_2h_2,s})\right| &\leq c_0,\\
\sum_{s=1}^{T}\sum_{l_2=1}^{p_1}\sum_{h_2=1}^{p_2}\left|Cov(e_{l_1j,t}e_{ih_1,t},e_{l_2j,s}e_{ih_2,s})\right| &\leq c_0.
\end{aligned}
$$

Setting $r=1$, it is easy to check by direct calculation that our version of Assumption A.3*(iii)(d)-(d')* implies all the relationships above (for example, the first one follows upon setting $j=k$; the second one with $i = h$; the third one with both $j = k$ and $i = h$; the fourth and the fifth instead follow directly). Indeed, upon inspecting the proofs in Yu et al. (2022) and He et al. (2023), these always require a bound on the growth of the second moment

$$E\left|\sum_{s=1}^{T}\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\left(e_{ij,t}e_{hk,t} - E\left(e_{ij,t}e_{hk,t}\right)\right)\right|^2,$$

so that our current version of Assumption A.3 may be viewed as (slightly) more primitive than the corresponding ones in Yu et al. (2022) and He et al. (2023).

Define now $\overline{\zeta}_{ij} = \mathrm{Vec}(T^{-1/2}\sum_{t=1}^{T}\mathbf{F}_t e_{ij,t})$.

Assumption A.4. *For some $r \geq 2$: (i) it holds that*

$$\begin{aligned}
\text{(a) } E\left\|\sum_{t=1}^{T}\mathbf{F}_t e_{ij,t}\right\|^{2r} &\leq c_0 T^r \text{ for all } 1 \leq i \leq p_1 \text{ and } 1 \leq j \leq p_2, \\
\text{(b) } E\left\|\sum_{t=1}^{T}\sum_{j=1}^{p_2}\mathbf{F}_t e_{ij,t}\right\|^{2r} &\leq c_0\left(Tp_2\right)^r \text{ for all } 1 \leq i \leq p_1, \\
\text{(c) } E\left\|\sum_{t=1}^{T}\sum_{i=1}^{p_1}\mathbf{F}_t e_{ij,t}\right\|^{2r} &\leq c_0\left(Tp_1\right)^r \text{ for all } 1 \leq j \leq p_2, \\
\text{(d) } E\left\|\sum_{t=1}^{T}\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\mathbf{F}_t e_{ij,t}\right\|^{2r} &\leq c_0\left(Tp_1p_2\right)^r;
\end{aligned}$$

(ii) it holds that

$$\begin{aligned}
\text{(a) } \left\|\sum_{h=1}^{p_2}E\left(\overline{\zeta}_{ij}\otimes\overline{\zeta}_{ih}\right)\right\|_{\max} &\leq c_0 \text{ for all } 1 \leq i \leq p_1 \text{ and } 1 \leq j \neq h_1 \leq p_2, \\
\text{(b) } \left\|\sum_{l=1}^{p_1}E\left(\overline{\zeta}_{ij}\otimes\overline{\zeta}_{lj}\right)\right\|_{\max} &\leq c_0 \text{ for all } 1 \leq i \neq l \leq p_1 \text{ and } 1 \leq j \leq p_2, \\
\text{(c) } E\left\|\sum_{i=1}^{p_1}\sum_{h=1}^{p_2}\left(\overline{\zeta}_{ih}\otimes\overline{\zeta}_{ik} - E\left(\overline{\zeta}_{ih}\otimes\overline{\zeta}_{ik}\right)\right)\right\|_{\max}^{r} &\leq c_0\left(p_1p_2\right)^{r/2} \text{ for all } 1 \leq j \leq p_1 \text{ and } 1 \leq k \leq p_2.
\end{aligned}$$

Assumption A.4 does not require independence between the common factors $\mathbf{F}_t$ and the errors $\mathbf{E}_t$, which are allowed to be (weakly) correlated. Indeed, whilst the assumption holds under the more restrictive case that $\{\mathbf{F}_t\}$ and $\{\mathbf{E}_t\}$ are two mutually independent groups, in He et al. (2023) some examples of dependence between $\{\mathbf{F}_t\}$ and $\{\mathbf{E}_t\}$ are considered. As far as part *(ii)(c)* of the assumption is concerned, we note that in Yu et al. (2022) and He et al. (2023), it is required that

$$\begin{aligned}
\left\|\sum_{i=1}^{p_1}\sum_{h=1}^{p_2}Cov\left(\overline{\zeta}_{ih}\otimes\overline{\zeta}_{ik},\overline{\zeta}_{i'h'}\otimes\overline{\zeta}_{i'k}\right)\right\|_{\max} &\leq c_0 \text{ for all } 1 \leq i' \leq p_1 \text{ and } 1 \leq h', k \leq p_2, \\
\left\|\sum_{i=1}^{p_1}\sum_{h=1}^{p_2}Cov\left(\overline{\zeta}_{ih}\otimes\overline{\zeta}_{jh},\overline{\zeta}_{i'h'}\otimes\overline{\zeta}_{jh'}\right)\right\|_{\max} &\leq c_0 \text{ for all } 1 \leq i', j \leq p_1 \text{ and } 1 \leq h' \leq p_2.
\end{aligned}$$

Both assumptions are implied by part *(ii)(c)* of the current version of the assumption, as can be easily verified by setting $r = 2$.

Our last assumption states that, under (3.2), the spaces spanned by the loadings before and after the break must differ.

ASSUMPTION A.5. *Under (3.2), it holds that both* $\mathbf{R}$ *and* $\widetilde{\mathbf{R}}$ *satisfy Assumption A.2 and, as* $p_1 \to \infty$

$$rank \begin{pmatrix} p_1^{-1}\mathbf{R}'\mathbf{R} & p_1^{-1}\mathbf{R}'\widetilde{\mathbf{R}} \\ p_1^{-1}\widetilde{\mathbf{R}}'\mathbf{R} & p_1^{-1}\widetilde{\mathbf{R}}'\widetilde{\mathbf{R}} \end{pmatrix} \geq k_1 + 1.$$

Assumption A.5 is a high-level condition to ensure that there is a genuine break in the loadings $\mathbf{R}$. In order to better understand the assumption, consider the following examples. We begin with the case, considered in Bai et al. (2022) in the context of vector-valued time series, where there is only a rotational change in the loading spaces

$$\begin{aligned} \mathbf{R} &= \overline{\mathbf{R}}\mathbf{A}, \\ \widetilde{\mathbf{R}} &= \overline{\mathbf{R}}\mathbf{B}, \end{aligned}$$

with $\overline{\mathbf{R}}$ a $p_1 \times k_1$ matrix with $rank\left(\overline{\mathbf{R}}\right) = k_1$, and $\mathbf{A}$ and $\mathbf{B}$ two $k_1 \times k_1$, full rank matrices. In such a case, (3.2) can be rewritten as

$$\mathbf{X}_t = \begin{cases} \overline{\mathbf{R}}\left(\mathbf{A}\mathbf{F}_t\right)\mathbf{C}' + \mathbf{E}_t = \overline{\mathbf{R}}\mathbf{F}_{1,t}\mathbf{C}' + \mathbf{E}_t & \text{for } 1 \leq t \leq m + t^* \\ \overline{\mathbf{R}}\left(\mathbf{B}\mathbf{F}_t\right)\mathbf{C}' + \mathbf{E}_t = \overline{\mathbf{R}}\mathbf{F}_{2,t}\mathbf{C}' + \mathbf{E}_t & \text{for } t > m + t^* \end{cases},$$

i.e. the change in the second order structure of $\mathbf{X}_t$ is due to heteroskedasticity in the common factors. As we mentioned in the introduction, our methodology is designed to be robust to this case, and to detect breaks only when these genuinely pertain to the row or column loadings. This may be viewed as desirable, since consistent estimation of the loadings space in the presence of heteroskedasticity in the common factors is still guaranteed, and a changepoint in the second moment of the common factors may be viewed as a second order problem (whereas, conversely, a changepoint in the loadings has more serious consequences; see Massacci, 2021, and Wang, 2023). Indeed, using for simplicity the normalisation $p_1^{-1}\overline{\mathbf{R}}'\overline{\mathbf{R}}=I_{k_1}$, it would follow that

$$\frac{1}{p_1}\begin{pmatrix} \mathbf{R}'\mathbf{R} & \mathbf{R}'\widetilde{\mathbf{R}} \\ \widetilde{\mathbf{R}}'\mathbf{R} & \widetilde{\mathbf{R}}'\widetilde{\mathbf{R}} \end{pmatrix} = \frac{1}{p_1}\begin{pmatrix} \mathbf{A}'\overline{\mathbf{R}}'\overline{\mathbf{R}}\mathbf{A} & \mathbf{A}'\overline{\mathbf{R}}'\overline{\mathbf{R}}\mathbf{B} \\ \mathbf{B}'\overline{\mathbf{R}}'\overline{\mathbf{R}}\mathbf{A} & \mathbf{B}'\overline{\mathbf{R}}'\overline{\mathbf{R}}\mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{A}'\mathbf{A} & \mathbf{A}'\mathbf{B} \\ \mathbf{B}'\mathbf{A} & \mathbf{B}'\mathbf{B} \end{pmatrix}.$$

Hence, the fact that this matrix has rank equal to k_1 follows from standard passages; indeed, since both $\mathbf{A}$ and $\mathbf{B}$ have full rank, there exists a full rank matrix $\mathbf{P}$ such that $\mathbf{AP} = \mathbf{B}$, so that

$$\begin{aligned} \begin{pmatrix} \mathbf{A}'\mathbf{A} & \mathbf{A}'\mathbf{B} \\ \mathbf{B}'\mathbf{A} & \mathbf{B}'\mathbf{B} \end{pmatrix} &= \begin{pmatrix} I_{k_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}' \end{pmatrix}\begin{pmatrix} \mathbf{A}' \\ \mathbf{A}' \end{pmatrix}\begin{pmatrix} \mathbf{A} & \mathbf{A}' \end{pmatrix}\begin{pmatrix} I_{k_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} \end{pmatrix} \\ &= \begin{pmatrix} I_{k_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}' \end{pmatrix}\left[\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \mathbf{A}'\mathbf{A}\right]\begin{pmatrix} I_{k_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} \end{pmatrix}. \end{aligned}$$

It is now easy to see that the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \mathbf{A}'\mathbf{A},$$

has rank k_1.

As a second example, consider the case where $\widetilde{\mathbf{R}}'\mathbf{R} = \mathbf{0}$. In such a case, it is immediate to see that, as $p_1 \to \infty$, the matrix

$$\begin{pmatrix} p_1^{-1}\mathbf{R}'\mathbf{R} & \mathbf{0} \\ \mathbf{0} & p_1^{-1}\widetilde{\mathbf{R}}'\widetilde{\mathbf{R}} \end{pmatrix},$$

has full rank $2k_1$, and therefore Assumption A.5 is satisfied.

B. Extensions, implementation guidelines, and further simulations. In this section, we consider several extensions that complement the theory spelt out in the main paper (Section B.1); we offer a more detailed discussion of our implementation guidelines (Section B.2); and we report a comprehensive set of simulations to assess the sensitivity of our procedures under different scenarios (Section B.3).

B.1. *Extensions.* We consider several extensions that complement the theory developed in the main paper: *(i)* the case of vanishing factors; *(ii)* the case of a one-way factor structure, where $k_1 = 0$ and/or $k_2 = 0$; *(iii)* the presence of weak factors; *(iv)* the case of the presence of changes in the column loadings matrix $\mathbf{C}$; *(v)* the approximation of critical values when using extreme value distributions; and, finally, *(vi)* the possible use of non-Gaussian sequences in our randomisation algorithm.

B.1.1. *Further changepoint alternatives: the case of vanishing factors .* Most of the focus of this paper has been on alternatives where the number of factors increases after t^*. One alternative hypothesis which is left out from this framework is the case where the column space of $\mathbf{R}$ reduces. In this case, (3.1) becomes

$$H_{A,3} : \mathbf{R} = \left[\widetilde{\mathbf{R}}|\mathbf{R}_3\right], \tag{B.1}$$

where $\mathbf{R}_3$ is a $p_1 \times c_4$ matrix, and then $\mathbf{F}'_{1,t} = \left[\mathbf{F}'_{2,t}|\widetilde{\mathbf{F}}'_{1,t}\right]$, and $\widetilde{\mathbf{F}}_{1,t}$ is a $c_4 \times k_2$ matrix for some $c_4 \geq 1$. By standard arguments, under (B.1) it holds that

$$\widehat{\lambda}_{k_1,\tau} \begin{cases} \geq c_0 p_1 & \text{for } \tau \leq t^* \\ \leq c_1 \left(1 - \frac{\tau - t^*}{m}\right) p_1 & \text{for } t^* < \tau < m + t^* \\ \leq c_0 l_{p_1,p_2,m} & \text{for } \tau \geq m + t^* \end{cases}. \tag{B.2}$$

In order to detect changes of the type (B.2), note that under the null of no change, it holds that

$$p_1^{-\delta}\widehat{\lambda}_{k_1,\tau} = \Omega_{a.s.}\left(p_1^{1-\delta}\right),$$

where δ is defined in (3.9), whereas in the presence of a break, by (B.2) it holds that there exists a $\widetilde{\tau} > t^*$ such that, for $\tau \geq \widetilde{\tau}$, $p_1^{-\delta}\widehat{\lambda}_{k_1,\tau} = o_{a.s.}(1)$. Define

$$\widetilde{\psi}_\tau = \left(\left|\frac{p_1^{-\delta}\widehat{\lambda}_{k_1,\tau}}{p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau}}\right|^r\right)^{-1}. \tag{B.3}$$

By continuity, it follows that

$$\widetilde{\psi}_\tau = o_{a.s.}(1) \text{ for all } 1 \leq \tau \leq T_m \text{ under the null of no changepoint,}$$

and

$$\widetilde{\psi}_\tau \stackrel{a.s.}{\to} \infty \text{ for some } \tau \geq \widetilde{\tau} > t^*, \text{ in the presence of a changepoint in } t^*.$$

From hereon, the procedures described in Section 3 can be applied, using $\widetilde{\psi}_\tau$ in lieu of ψ_τ.

PROPOSITION A.1. *We assume that the assumptions of Theorems 1-3 hold, and that*

$$\lim_{\min\{p_1,m\}\to\infty} \frac{m}{p_1^{r(1-\delta)}} \to 0. \tag{B.4}$$

Then, the same conclusions as in Theorems 1-3 hold for $\widetilde{\psi}_\tau$.

Proposition A.1 states that our monitoring schemes can be used with $\widetilde{\psi}_{\tau}$. Under the null, the same procedure-wise size control obtains, as when using ψ_{τ}.

Under the alternative, results are more delicate and we discuss them heuristically hereafter. In the construction of the monitoring procedure, there are some specifications to be decided. As we note in Section 4.1, r is given, and the tuning parameters are m and δ. Then, similarly to Section 4.1, δ can be fixed in advance and m chosen thereafter, to satisfy (B.4) for the purpose of size control, and to minimise the detection delay under the alternative. Under (B.1), equation (B.2) entails that in the presence of a changepoint at time t^*+1, the k_1-th eigenvalue calculated at $\tau \geq t^*+1$ - denoted as $\widehat{\lambda}_{k_1,\tau}$ - will be proportional to $\frac{\tau-t^*}{m}p_1$. In the construction of our test statistics, we premultiply $\widehat{\lambda}_{k_1,\tau}$ by $p_1^{-\delta}$; in this case, this is done in order to ensure that, when there is no break, $p_1^{-\delta}\widehat{\lambda}_{k_1,\tau}$ diverges. Intuitively, this requires that the role played by δ is reversed, and that power is enhanced by larger values of δ, which is natural since $\widetilde{\psi}_{\tau}$ is constructed as the reciprocal of ψ_{τ}. In the presence of a changepoint, by (B.2) the sequence $\widetilde{\psi}_{\tau}$ is proportional to

$$p_1^{\delta-1}\frac{m}{m+\tau-t^*},$$

whenever $t^* < \tau < m+t^*$, which drifts to zero. The partial sum process $S_{\tau} = \sum_{j=0}^{\tau}\widetilde{\psi}_j + \sum_{j=0}^{\tau} z_j$ therefore has a non-centrality term proportional to $\sum_{j=t^*+1}^{\tau}\widetilde{\psi}_j$.

$$\begin{aligned}\sum_{j=t^*+1}^{\tau}\widetilde{\psi}_j &\sim c_0 p_1^{r(\delta-1)} m^r \sum_{h=1}^{\tau-t^*}\left(\frac{1}{m-h}\right)^r. \\ &\geq c_1 p_1^{r(\delta-1)}\left(\tau-t^*\right).\end{aligned}$$

Such a term can be split into two terms, viz.

$$\sum_{j=t^*+1}^{\tau}\widetilde{\psi}_j = \sum_{j=t^*+1}^{m-c_0+t^*}\widetilde{\psi}_j + \sum_{j=m-c_0+t^*+1}^{m}\widetilde{\psi}_j. \tag{B.5}$$

The first term is bounded by

$$\sum_{j=t^*+1}^{m-c_0+t^*}\widetilde{\psi}_j \sim c_0 p_1^{r(\delta-1)} m^r \sum_{h=1}^{m-c_0}\left(\frac{1}{m-h}\right)^r \leq c_1 m p_1^{r(\delta-1)} \to 0,$$

on account of (B.4) - hence, this term does not contribute to the detection delay. Conversely, as far as the second term in (B.5) is concerned, it holds that

$$\sum_{j=m-c_0+t^*+1}^{m}\widetilde{\psi}_j \sim c_0 p_1^{r(\delta-1)} m^r \sum_{h=m-c_0+1}^{m}\left(\frac{1}{m-h}\right)^r \geq c_1 p_1^{r(\delta-1)} m^r.$$

This term can pass to positive infinity, thus ensuring detection, as long as m is chosen such that $p_1^{(\delta-1)}m \to \infty$; however, the passages above entail that the detection delay is proportional to m. These heuristic considerations suggest that our methodology, based on exploiting the divergence rate of eigenvalues, is (very) effective at shortening the detection delay when the factor space enlarges (as in the cases treated in Section 3); conversely, when the factor space contracts, having $p_1 \to \infty$ has an adverse impact because it "hides" the fact that the k_1-th eigenvalue has become bounded: the blessing of dimensionality mentioned at the end of Section 3.1.1, here has turned into a curse of dimensionality.

We point out that ψ_τ and $\widetilde{\psi}_\tau$ can be used together, in order to design an "omnibus" methodology which is able to detect changepoints of the types (3.2)-(3.4) and (B.1) simultaneously, by modifying (3.11) as

$$y_\tau = z_\tau + \left(\psi_\tau + \widetilde{\psi}_\tau\right).$$

This procedure can be further improved by considering more than one eigenvalue, i.e. by using

$$\overline{\psi}_\tau = \left|\frac{p_1^{-\delta}\sum_{j=1}^{C}\widehat{\lambda}_{k_1+j,\tau}}{p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau}}\right|^r \quad \text{and} \quad \overline{\widetilde{\psi}}_\tau = \left(\left|\frac{p_1^{-\delta}\sum_{j=0}^{C}\widehat{\lambda}_{k_1-j,\tau}}{p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau}}\right|^r\right)^{-1},$$

for some user-chosen $1 \leq C < \infty$, and then employ our monitoring schemes with $y_\tau = z_\tau + \left(\overline{\psi}_\tau + \overline{\widetilde{\psi}}_\tau\right)$.[1]

In conclusion, we consider an alternative approach based on the eigenvalue ratio principle, using

$$\psi_\tau^* = m^{a(r/2-1)}\left|1 - \frac{\widehat{\lambda}_{k_1,\tau}}{\widehat{\lambda}_{k_1,0}}\right|^r, \tag{B.6}$$

for some user-chosen $a < 1$, in our monitoring scheme. Under the null, $\widehat{\lambda}_{k_1,\tau}$ and $\widehat{\lambda}_{k_1,0}$ are of comparable magnitude, and therefore ψ_τ^* drifts to zero; conversely, by (B.2), ψ_τ^* diverges at a rate $\Omega\left(m^{a(r/2-1)}\right)$.

Theorem A.1. *We assume that the assumptions of Theorems 1-3 hold with $r > 2$, and that*

$$\lim_{\min\{m,p_1,p_2\}\to\infty} \frac{m^{1+a(r/2-1)}}{\min\{p_1^r, p_2^r\}} \to 0. \tag{B.7}$$

Then, the same conclusions as in Theorems 1-3 hold for ψ_τ^*.

In the proof, we show that the partial sums have a non-centrality term given by

$$m^{a(r/2-1)}\sum_{\tau=t^*+1}^{T_m}\left|\frac{\tau - t^*}{m}\right|^r = c_0 m^{a(r/2-1)-r}\left(\tau - t^*\right)^{r+1}.$$

This entails that the detection delay is of order $O\left(m^{(r-a(r/2-1))/(r+1)}\right)$, i.e. smaller than $O(m)$ but larger than $\Omega\left(m^{1/2}\right)$ for all choices of a. This indicates a possible improvement compared to using $\widetilde{\psi}_\tau$; however, a low-dimensional approach like the one proposed by Bai et al. (2022) are likely to yield better results, at least in theory. The choice of a is a matter of tuning, and it reflects the trade-off between size and power: the larger a, the shorter the detection delay, and vice versa. Similarly to the guidelines in Section 4.1, a possible strategy is to fix a; then, given a, m can be chosen based on (B.7) as

$$m = o\left(\left(\min\{p_1, p_2\}\right)^{r/(1+a(r/2-1))}\right).$$

[1]We are grateful to an anonymous Referee for suggesting this to us.

In both cases (using $\widetilde{\psi}_\tau$ and ψ^*_τ), the theory spelt out above suggests that the choice of m in this case is important, and that a shorter delay can be obtained with a smaller m.

We conclude this section by reporting some Monte Carlo evidence. Using the same design as in Section 4.2 in the main paper, we consider the performance of methodologies based on using both $\widetilde{\psi}_\tau$ and ψ^*_τ under

$$H_{A,3} : \mathbf{R} = \left[\widetilde{\mathbf{R}}|\mathbf{R}_3\right],$$

where $\mathbf{R}_3$ is a $p_1 \times 1$ vector, and $\mathbf{F}'_{1,t} = \left[\mathbf{F}'_{2,t}|\widetilde{\mathbf{F}}'_{1,t}\right]$, with $\widetilde{\mathbf{F}}_{1,t}$ a $1 \times k_2$ vector - that is, only one factor vanishes. When using ψ^*_τ, we have set $a = 0.3$; unreported simulations show that if a increases, this leads to better power but worse size control, as can be expected - by way of a guideline, we thus recommend using $a = 0.3$, which resulted in the best balance between power and size in our experiments. Results are in Table B.1 - as can be seen, methods based on ψ^*_τ deliver a better detection performance. In general, both methodologies based on $\widetilde{\psi}_\tau$ and ψ^*_τ require p_2 to be larger than 20 for size control. As predicted by the theory, we note that delays tend to be longer than in the case of new factors emerging; interestingly, the detection delay worsens as p_2 increases, because our adaptive rules for the selection of m allow for a larger m the larger p_2. In principle, it would be possible to determine an "optimal" value of m, but as mentioned above, in this case, a methodology based on a different paradigm that does not use the k_1-th largest eigenvalue may be preferable.

B.1.2. *The cases $k_1 = 0$ and $k_2 = 0$.* In the main paper, we have considered the case where both $k_1 > 0$ and $k_2 > 0$: this is the genuine matrix factor model, where, according to (1.1), the matrix structure of the data comes into play. However, it is entirely possible that either $k_1 = 0$, or $k_2 = 0$, or both during the training sample: for example, it is possible that there is no row factor structure in $1 \leq t \leq m$, but then a factor structure arises, after a changepoint, among the rows of the dataset. In this section, we briefly discuss these cases, and how to modify our approach. Similarly to He et al. (2023), we use the following CONVENTION to characterize such boundary cases

$$\underset{p_1\times p_2}{\mathbf{X}_t} = \begin{cases} \underset{p_1\times k_1}{\mathbf{R}}\underset{k_1\times p_2}{\mathbf{F}_t} + \underset{p_1\times p_2}{\mathbf{E}_t} & \text{when } k_1 > 0, k_2 = 0, \\ \underset{p_1\times k_2}{\mathbf{F}_t}\underset{k_2\times p_2}{\mathbf{C}'} + \underset{p_1\times p_2}{\mathbf{E}_t} & \text{when } k_2 > 0, k_1 = 0, \\ \underset{p_1\times p_2}{\mathbf{E}_t} & \text{when } k_1 = k_2 = 0, \end{cases} \tag{B.8}$$

where the first case refers to a one-way factor model along the row dimension (all columns form a vector factor model), the second case is a one-way factor model along the column dimension (all rows form a vector factor model), and the third case means absence of any factor structure.

We begin by discussing the (easier) case $k_1 > 0$ and $k_2 = 0$. In such a case, there is no scope to use $\widehat{\mathbf{M}}_1$ and the spectrum thereof. Hence, upon finding that $k_2 = 0$ (using e.g. the tests discussed in He et al., 2023), the eigenvalues of $\mathbf{M}_r$ defined in (2.2) can be used instead. In particular, letting $\widetilde{\lambda}_j$ denote the j-th largest eigenvalue of $\mathbf{M}_r$, in Theorem 1 in He et al. (2023), it is shown that, whenever $k_1 > 0$ and $k_2 = 0$

$$\widetilde{\lambda}_j = \Omega(p_1) \text{ for } j \leq k_1, \tag{B.9}$$

$$\widetilde{\lambda}_j = c_0 + o_{a.s.}(l_{p_1,p_2,m}) \text{ for } j > k_1, \tag{B.10}$$

and therefore $\widetilde{\lambda}_{k_1+1,\tau}$ - computed across the monitoring horizon $1 \leq \tau \leq T_m$ - can be used in the same way as $\widehat{\lambda}_{k_1+1,\tau}$: all our theory can be adapted to this case, indeed being easier to do so since the estimator $\widetilde{\mathbf{C}}$ is not required.

Table B.1

Results when the last row factor vanishes, over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\epsilon = 0.05$ and $T = 200$. "PS", partial-sum; "WC", worst case.

			$\widetilde{\psi}_\tau$				ψ^*_τ			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical size ($\times 100$) under H_0										
200	20	80	4.8	1.6	3.0	3.4	4.1	2.3	3.1	3.4
200	30	80	5.0	1.6	3.2	4.0	4.4	2.2	3.7	5.0
200	40	80	4.3	1.7	2.9	4.3	4.1	2.1	3.0	4.3
200	80	20	3.8	1.8	3.2	5.1	12.4	9.4	8.3	11.5
200	80	30	5.0	2.6	3.3	4.9	6.3	4.1	4.8	7.2
200	80	40	4.4	1.9	2.6	4.2	5.2	3.3	4.7	5.2
200	20	20	5.5	3.8	3.8	4.8	11.2	8.6	7.6	12.5
200	30	30	5.2	2.4	4.2	3.1	6.6	3.2	4.7	6.9
200	40	40	4.4	2.0	3.4	3.5	4.5	2.6	4.2	5.0
Empirical power ($\times 100$) when a row factor vanishes										
200	20	80	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	30	80	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	40	80	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	80	20	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	80	30	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	80	40	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	20	20	99.8	99.7	98.0	99.2	90.0	84.2	66.7	66.3
200	30	30	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	40	40	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Empirical delay ($\times 100$) when a row factor vanishes										
200	20	80	69,72,74	69,72,75	71,74,76	67,70,73	62,67,71	64,68,72	68,71,75	60,64,68
200	30	80	74,75,77	74,76,77	75,77,78	72,74,76	61,65,69	62,66,70	66,70,73	59,63,67
200	40	80	75,77,78	76,77,78	77,78,79	74,75,77	60,64,68	62,65,69	65,69,72	58,62,65
200	80	20	16,18,18	16,18,19	18,19,20	15,17,18	19,21,22	20,21,23	23,25,26	17,18,19
200	80	30	27,28,29	27,28,29	28,29,30	26,27,28	27,28,29	27,29,30	30,31,32	24,26,27
200	80	40	37,38,39	37,38,39	39,39,40	37,38,38	34,36,37	35,36,38	37,39,40	31,33,35
200	20	20	19,22,27	20,23,30	22,26,38	17,19,21	27,35,48	29,38,54	37,49,66	19,25,45
200	30	30	28,29,30	29,30,31	30,31,33	27,28,29	28,30,31	29,30,32	32,34,36	25,27,28
200	40	40	38,39,39	38,39,40	39,40,41	37,38,39	35,36,38	35,37,39	38,40,41	32,34,35

When $k_1 = 0$, upon assuming, in addition to Assumptions A.1-A.4, that

$$\lambda_{\max}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t'\right) = O_{a.s.}\left(\left(1+\sqrt{\frac{p_1}{T}}\right)^2\right), \tag{B.11}$$

He et al. (2023) show the following result, which we state as a lemma

Lemma A.1. *We assume that Assumptions A.1-A.4 and (B.11) hold. Then, for all $j > 0$*

$$\begin{aligned}\widehat{\lambda}_j &= O_{a.s.}\left(\frac{p_1}{T}\right) + o_{a.s.}\left(l_{p_1,p_2,m}\right),\\ \widetilde{\lambda}_j &= c_0 + O_{a.s.}\left(\frac{p_1}{T}\right) + o_{a.s.}\left(l_{p_1,p_2,m}\right).\end{aligned}$$

Based on this, upon modifying (3.9) as

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

with $\beta' = \ln p_1 / \min\{\ln(p_2 m), 2\ln(m)\}$, it is possible to show (by repeating and marginally adapting our proofs) that all the results derived in the main paper still hold. Finally, we point out that the procedure to estimate k_1 and k_2 reported in Section B.2.1 can also be used to to check whether $k_1 > 0$ and $k_2 > 0$, thus making our procedure fully feasible.

In Table B.2, we report some Monte Carlo evidence on the case $k_1 = 0$. Results are broadly good, at least for the large sample case.

Table B.2

Results when $k_i = 0$, over 1000 replications, with significance level $\alpha = 0.05$, $m = p_2$, $r = 8$, $\epsilon = 0.05$. "PS", partial-sum; "WC", worst case.

			$k_1 = 3$, $k_2 = 0$ (without projection)				$k_1 = 0$, $k_2 = 3$ (with projection)			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical size ($\times 100$) under H_0										
200	80	20	2.9	2.1	2.2	5.5	3.9	1.2	3.5	3.5
200	80	30	4.5	2.2	3.7	3.1	5	1.1	3.7	3.1
200	80	40	5.3	1.7	3.5	4.2	3.8	1.5	3	5.3
Empirical power ($\times 100$) under H_{A1}										
200	80	20	5.4	2.9	3.6	5.4	100	100	100	100
200	80	30	92.9	89	76.3	96	100	100	100	100
200	80	40	100	100	100	100	100	100	100	100
Empirical delay (interquantiles) under H_{A1}										
200	80	20	9,28,78	-70,-47,2	-74,-74,-72	-5,10,56	7,8,10	7,8,10	8,10,11	5,7,8
200	80	30	7,7,9	7,8,9	8,9,11	5,6,6	5,7,8	5,7,8	6,8,9	4,5,7
200	80	40	6,6,7	6,6,7	7,8,8	5,5,6	5,6,7	5,6,7	6,7,8	3,5,6

B.1.3. *Changepoint detection in the presence of weak factors.* Assumption A.2 states that $\|\mathbf{R}\|^2 = c_0 p_1$; in turn, this entails that the (row) common factors $\mathbf{F}_t$ are "strong" or "pervasive" (see e.g. the recent contributions by Uematsu and Yamagata (2021) for a thorough analysis of this case in the vector factor model case). By standard arguments, it is easy to see that this corresponds to the case where the spiked eigenvalues of $\widehat{\mathbf{M}}_1$ diverge as fast as p_1.

In this section, we study the case of weak factors, where $\|\mathbf{R}\|^2 = c_0 p_1^{\alpha_R}$ for some $0 < \alpha_R < 1$,[2] under the maintained assumption that $\|\mathbf{C}\|^2 = c_0 p_2$, i.e. that the column factors are strong. We would like to point out that it is entirely possible to consider also the case of weak column factors, i.e. $\|\mathbf{C}\|^2 = c_0 p_2^{\alpha_C}$, with $0 < \alpha_C < 1$ - indeed, in the context of estimating k_1 and k_2, He et al. (2023) study this extension, which we do not report in order to not overshadow the main results with the algebra.

We consider only one leading example, for the sake of a concise discussion, where, under the alternative (3.4), a new (weak) factor arises, i.e.

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

where recall that $\widetilde{\mathbf{R}} = [\mathbf{R}|\mathbf{R}_3]$, $\mathbf{R}_3$ is a $p_1 \times 1$ matrix, $\mathbf{F}_{2,t}' = \left[\mathbf{F}_{1,t}' | \widetilde{\mathbf{F}}_{2,t}'\right]$, and $\widetilde{\mathbf{F}}_{2,t}$ is a $1 \times k_2$ new common factor. We assume that the new loadings are orthogonal to the previous ones, and that $\widetilde{\mathbf{F}}_{2,t}$ is indeed a weak factor, viz.

$$\mathbf{R}'\mathbf{R}_3 = 0 \ \text{ and } \ p_1^{-\alpha}\mathbf{R}_3'\mathbf{R}_3 \to 1, \tag{B.12}$$

as $p_1 \to \infty$. Whilst this is only a toy example, it allows a straightforward analysis of the impact of weak factors on the detection delay. Upon repeating the proof of Lemma C.3*(ii)*, it is easy to see that in this case

$$\widehat{\lambda}_{k_1+1,\tau} \begin{cases} \le c_0 & \text{for } \tau \le t^* \\ \ge c_1 \frac{\tau - t^*}{m} p_1^{\alpha_R} & \text{for } t^* < \tau < m + t^* \\ \ge c_0 p_1^{\alpha_R} & \text{for } \tau \ge m + t^* \end{cases}.$$

Repeating the proof of Theorems 2 and 3, it can be shown that our procedures have power versus this alternative as long as

$$\lim_{p_1 \to \infty} p_1^{\alpha_R - \delta} = \infty. \tag{B.13}$$

In turn, this can be further analysed in the light of (3.9). Equation (B.13) is satisfied as long as $\alpha_R > \delta$. Whenever $\beta \le 1/2$, i.e. whenever

$$\frac{p_1}{\sqrt{m p_2}} \to 0,$$

(B.13) will always hold - this means that when p_1 is not "too large" compared to $m p_2$, breaks where a new weak factor emerge can be detected even when α_R is very small. This can be compared with the case of vector factor models, where detection requires $p_1/\sqrt{m} \to 0$, thus being more restrictive (see Barigozzi and Trapani, 2020). Conversely, when $\beta > 1/2$, (B.13) requires

$$\alpha_R > 1 - \frac{1}{2}\frac{\ln m p_2}{\ln p_1},$$

[2]We are grateful to an anonymous Referee for asking the question that led to this section.

which indicates that breaks involving the emerging of a new, weak factor can be detected as long as the new factor is not "too weak". For example, if $p_1 = mp_2$, then this entails that $\alpha_R > 1/2$, otherwise new weak factors cannot be detected. These considerations have an immediate impact on the detection delay: the sufficient condition in (3.27) now becomes less likely to hold, viz.

$$\frac{p_1^{\alpha_R-\delta}}{m}(\tau - t^*) \to \infty,$$

which indicates that the delay will be larger when breaks involve weak factors.

As a final remark, we briefly consider the alternative (3.3) where one factor becomes weak after a changepoint t^*

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

where $\mathbf{R} = [\mathbf{R}_0|\mathbf{R}_1]$, $\widetilde{\mathbf{R}} = [\mathbf{R}_0|\mathbf{R}_2]$, $\mathbf{R}_0$ is a $p_1 \times (k_1 - 1)$ matrix of loadings which do not undergo a change, and $\mathbf{R}_1$ and $\mathbf{R}_2$ are $p_1 \times 1$ vectors of loadings which differ before and after the changepoint t^*. In such a case, following the proof of Lemma C.3*(i)*, it can be shown that

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

and the same logic as above can be used to study this case too.

In Table B.3, we evaluate the performance of our methodology in the presence of one weak factor. We consider two rescalings: one, suggested in the paper, based on the average of the eigenvalues of the projected matrix, and one based on the average of the eigenvalues of the non projected matrix, i.e.

$$\frac{p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}}{p_1^{-1}\sum_{j=1}^{p_1}\widetilde{\lambda}_{j,\tau}}.$$

In the former case, we note that size control is more challenging when α_R is small; conversely, better size control (at the price of lower power) is achieved with the latter rescaling. Indeed, this can be expected. When $p_1^{-\alpha_R}R'R \to I_{k_1}$ for some $\alpha_R \in (0,1)$, after projection it can be expected that $\widehat{\lambda}_j \asymp p_1^{\alpha_R}$ for $1 \leq j \leq k_1$, while $p_1^{-1}\sum_{j=k_1+1}^{p_1}\widehat{\lambda}_j \asymp p_2^{-1}$. Therefore, the denominator of (3.10) will be roughly $p_1^{\alpha_R-1} + p_2^{-1}$, which converges to 0, whence the loss of size control. Conversely, when rescaling by the eigenvalues of the non-projected second moment matrix, $\widetilde{\lambda}_j$, it can be shown (following the arguments in He et al., 2023) that the denominator is roughly a constant.

B.1.4. *Changes in the column factors loadings.* In the main paper, we focus on the case where $\mathbf{C}$ is constant over time. This simplifies the presentation, but it may be considered an unrealistic set-up. We now provide some arguments which show that our procedure is able to detect changes in $\mathbf{R}$ also in the possible presence of changes in $\mathbf{C}$. Inspired by Baltagi et al. (2017), we consider the following scenario

$$\mathbf{X}_t = \begin{cases} \mathbf{R}\mathbf{F}_{1,t}\mathbf{C}_t' + \mathbf{E}_t & \text{for } 1 \leq t \leq m + t^* \\ \widetilde{\mathbf{R}}\mathbf{F}_{2,t}\mathbf{C}_t' + \mathbf{E}_t & \text{for } t > m + t^* \end{cases}, \tag{B.14}$$

where we allow for

$$\mathbf{C}_t = \begin{cases} \mathbf{C}_1 & \text{for } 1 \leq t \leq m + \widetilde{t} \\ \mathbf{C}_2 & \text{for } t > m + \widetilde{t} \end{cases},$$

Table B.3
Results with weak factors for the projected method, over 1000 replications, with significance level 0.05*,* $k1 = k_2 = 3$*,* $m = p_2$*,* $r = 8$*,* $\epsilon = 0.05$*, and* $T = 200$*. "PS", partial-sum; "WC", worst case.*

			Rescale with $\hat{\lambda}_{j,\tau}$				Rescale with $\tilde{\lambda}_{j,\tau}$			
			PS			WC	PS			WC
p_1	p_2	α	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical power (×100) under H_{A3}										
40	40	1	4.3	1.9	2.9	3.0	3.8	1.9	4.0	3.2
40	40	0.8	12.2	6.6	5.4	6.3	4.7	2.4	3.4	5.0
40	40	0.6	99.8	99.5	97.6	84.2	4.7	2.1	2.7	5.0
40	40	0.4	100.0	100.0	100.0	100.0	3.9	1.9	2.7	3.8
60	60	0.4	100.0	100.0	100.0	100.0	3.7	1.5	3.4	3.2
80	80	0.4	100.0	100.0	100.0	100.0	4.6	1.5	2.7	5.0
Empirical power (×100) under H_{A1}										
40	40	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
40	40	0.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
40	40	0.6	100.0	100.0	100.0	100.0	98.6	97.6	95.6	95.9
40	40	0.4	100.0	100.0	100.0	100.0	35.5	23.3	9.9	15.6
60	60	0.4	100.0	100.0	100.0	100.0	66.7	52.9	32.0	38.7
80	80	0.4	100.0	100.0	100.0	100.0	87.2	80.2	59.8	68.3
Empirical delay (interquantiles) under H_{A1}										
40	40	1	6,7,9	6,7,9	7,9,10	5,6,7	10,12,14	10,12,14	11,13,16	8,10,12
40	40	0.8	6,7,9	6,8,9	7,9,11	5,6,8	14,17,20	14,17,21	16,19,23	11,14,18
40	40	0.6	-53,-46,-34	-57,-52,-40	-54,-54,-43	-53,-30,5	24,29,35	25,30,37	28,33,41	21,25,32
40	40	0.4	-59,-59,-58	-59,-59,-59	-54,-54,-54	-59,-59,-59	51,66,82	50,67,85	-52,60,80	39,55,74
60	60	0.4	-39,-38,-37	-39,-39,-39	-35,-35,-35	-39,-39,-39	56,66,81	57,68,81	61,74,88	53,62,79
80	80	0.4	-18,-17,-16	-19,-19,-18	-15,-15,-15	-19,-19,-18	60,68,78	62,72,82	67,77,86	59,68,77

and - only for simplicity and for illustrative purposes - we assume that: $\widetilde{t} < t^*$, $\mathbf{C}_1$ and $\mathbf{C}_2$ have the same number of columns k_2 and full rank. Define now the (full rank) $p_2 \times k_3$ matrix $\mathbf{C}_3$ (where $k_3 \geq k_2$) such that

$$\mathbf{C}_1 = \mathbf{C}_3 \mathbf{\Omega}_1 \quad \text{and} \quad \mathbf{C}_2 = \mathbf{C}_3 \mathbf{\Omega}_2,$$

where $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are $k_3 \times k_2$ matrices with full rank k_2. The equation above stipulates that the column spaces of both $\mathbf{C}_1$ and $\mathbf{C}_2$ are spanned by the column space of $\mathbf{C}_3$. In particular, if $\mathbf{C}_1$ and $\mathbf{C}_2$ are orthogonal, then $k_3 = 2k_2$; if they both lie in the same colum space, $k_3 = k_2$; and, in general, $k_2 \leq k_3 \leq 2k_2$. Hence we can rewrite (B.14) as

$$\mathbf{X}_t = \begin{cases} \mathbf{R}\mathbf{F}_{1,t}\mathbf{\Omega}_1'\mathbf{C}_3' + \mathbf{E}_t & \text{for } 1 \leq t \leq m + \widetilde{t} \\ \mathbf{R}\mathbf{F}_{1,t}\mathbf{\Omega}_2'\mathbf{C}_3' + \mathbf{E}_t & \text{for } m + \widetilde{t} < t \leq m + t^* \\ \widetilde{\mathbf{R}}\mathbf{F}_{2,t}\mathbf{\Omega}_2'\mathbf{C}_3' + \mathbf{E}_t & \text{for } t > m + t^* \end{cases} .$$

Finally, letting

$$\widetilde{\mathbf{F}}_t = \begin{cases} \mathbf{F}_t\mathbf{\Omega}_1' & \text{for } 1 \leq t \leq m + \widetilde{t} \\ \mathbf{F}_t\mathbf{\Omega}_2' & \text{for } m + \widetilde{t} < t \leq m + t^* \\ \mathbf{F}_t\mathbf{\Omega}_2' & \text{for } t > m + t^* \end{cases} ,$$

we can write

$$\mathbf{X}_t = \begin{cases} \mathbf{R}\widetilde{\mathbf{F}}_t\mathbf{C}_3' + \mathbf{E}_t & \text{for } 1 \leq t \leq m + t^* \\ \widetilde{\mathbf{R}}\widetilde{\mathbf{F}}_t\mathbf{C}_3' + \mathbf{E}_t & \text{for } t > m + t^* \end{cases} . \tag{B.15}$$

Upon extending Assumption A.1*(ii)* by requiring that, as $m \to \infty$

$$\frac{1}{m}\sum_{t=1}^{m} \mathbf{F}_{i,t}\mathbf{A}\mathbf{F}_{i,t}' \xrightarrow{P} \widetilde{\mathbf{\Sigma}}_i,$$

with $\widetilde{\mathbf{\Sigma}}_i$ positive definite for $i = 1, 2$ and all positive definite matrices $\mathbf{A}$, it is possible to use the theory developed in Yu et al. (2022) to estimate $\mathbf{C}_3$, thereby obtaining the same results as above. In Section B.3.5, we complement the theory developed above with a small Monte Carlo exercise.

B.1.5. *Approximations to the extreme value distribution.* In this section, we study how to refine the extreme value limiting distribution obtained in Section 3.

Approximation of (3.21). We begin with a theorem which provides a rate of the Gaussian approximation for partial sums. Recall the definition of S_u in (3.13)

$$S_u = \sum_{j=0}^{u} y_j,$$

and consider the sequence

$$a_m = \exp\left((\ln(T_m))^{1-\epsilon}\right),$$

where $\epsilon > 0$ is understood to be "arbitrarily small". Then it holds that

THEOREM A.2. *We assume that Assumptions A.1-A.4 and 1-3 hold, and that (3.12) holds. Then, under the null, as* $\min\{p_1, p_2, m\} \to \infty$, *on a suitably enlarged probability space, there exists a standard Wiener process* $\{W_{T_m}(\tau), 1 \le \tau \le T_m\}$ *such that*

$$\sup_{\frac{a_m}{T_m} \le u \le 1} \left| \frac{|S_u|}{u^{1/2}} - \frac{|W_{T_m}(u)|}{u^{1/2}} \right| = O_{P^*}\left(\frac{(\ln(T_m))^{1-\epsilon}}{\exp\left(\frac{1}{2}(\ln(T_m))^{1-\epsilon}\right)} \right),$$

for almost all realisations of $\{\mathbf{X}_t, 1 \le t \le T\}$.

The approximation in Theorem A.2 suggests that critical values for

$$\max_{1 \le \tau \le T_m} \frac{|S_\tau|}{\tau^{1/2}},$$

can be approximated using $\sup_{\frac{a_m}{T_m} \le u \le 1} u^{-1/2} |W_{T_m}(u)|$. The distribution of $\{W_{T_m}(\tau), 1 \le \tau \le T_m\}$ does not depend on T_m; thus

$$\sup_{\frac{a_m}{T_m} \le u \le 1} \frac{|W_{T_m}(u)|}{u^{1/2}} \stackrel{D}{=} \sup_{\frac{a_m}{T_m} \le u \le 1} \frac{|W(u)|}{u^{1/2}}.$$

Using the scale transformation of the Wiener process, it holds that

$$\sup_{\frac{a_m}{T_m} \le u \le 1} \frac{|W(u)|}{u^{1/2}} \stackrel{D}{=} \sup_{1 \le s \le \frac{T_m}{a_m}} \frac{|W(s)|}{s^{1/2}} \stackrel{D}{=} \sup_{1 \le s \le \exp\left(\ln\left(\frac{T_m}{a_m}\right)\right)} \frac{|W(s)|}{s^{1/2}}.$$

Using equation (18) in Vostrikova (1982), it holds that, as $x \to \infty$

$$P\left(\sup_{1 \le s \le \exp\left(\ln\left(\frac{T_m}{a_m}\right)\right)} \frac{|W(s)|}{s^{1/2}} > x \right) \simeq \frac{x \exp\left(-\frac{1}{2}x^2\right)}{\sqrt{2\pi}} \left(\ln\left(\frac{T_m}{a_m}\right) - \frac{1}{x^2} \ln\left(\frac{T_m}{a_m}\right) + \frac{4}{x^2} + O\left(x^{-4}\right) \right).$$

Hence, instead of $c_{\alpha,m}$ defined in (3.21), an approximate critical value, say $\widetilde{c}_{\alpha,m}$ can be used, defined as the solution to

$$\frac{\widetilde{c}_{\alpha,m} \exp\left(-\frac{1}{2}\widetilde{c}^2_{\alpha,m}\right)}{\sqrt{2\pi}} \left(\ln\left(\frac{T_m}{a_m}\right) - \frac{1}{\widetilde{c}^2_{\alpha,m}} \ln\left(\frac{T_m}{a_m}\right) + \frac{4}{\widetilde{c}^2_{\alpha,m}} \right) = \alpha.$$

Given that $a_m = o(T_m)$, this is tantamount to defining $\widetilde{c}_{\alpha,m}$ as the solution to

$$\frac{\widetilde{c}_{\alpha,m} \exp\left(-\frac{1}{2}\widetilde{c}^2_{\alpha,m}\right)}{\sqrt{2\pi}} \left(\ln(T_m) - \frac{1}{\widetilde{c}^2_{\alpha,m}} \ln(T_m) + \frac{4}{\widetilde{c}^2_{\alpha,m}} \right) = \alpha. \tag{B.16}$$

(Penultimate) approximation of (3.32). Upon inspecting the proof of Theorem 3, it follows that under the null, as $\min\{p_1, p_2, m\} \to \infty$

$$P^*\left(\frac{Z_{T_m} - b_{T_m}}{a_{T_m}} \le v \right) = P\left(\frac{Z^*_{T_m} - b_{T_m}}{a_{T_m}} \le v \right) + o_{P^*}(1),$$

where

$$Z^*_{T_m} = \max_{1 \le \tau \le T_m} z_\tau,$$

and $z_\tau \sim i.i.d.N(0,1)$. Thus, critical values for Z_{T_m} are based on $P\left(Z^*_{T_m} \leq a_{T_m}v + b_{T_m}\right)$. Theorem 3 suggests using the asymptotic critical values based on the Gumbel distribution; as an alternative, critical values based on the so-called "penultimate approximation" can be used (Gomes and Haan, 1999), based on approximating $P\left(Z^*_{T_m} \leq a_{T_m}v + b_{T_m}\right)$ with a Weibull distribution, viz.

$$P\left(\frac{Z_{T_m} - b_{T_m}}{a_{T_m}} \leq v\right) = \exp\left(-\left(1 + \gamma_{T_m} v\right)^{-1/\gamma_{T_m}}\right),$$

where $\gamma_{T_m} = -\left(2 \ln T_m\right)^{-1}$. Hence, after some algebra, the critical value $c_{\alpha,2}$ defined in (3.32) can be approximated by

$$\widetilde{c}_{\alpha,2} = b_{T_m} - a_{T_m}\left(-\frac{1}{\gamma_{T_m}} + \frac{1}{\gamma_{T_m}}\left(-\ln\left(1-\alpha\right)\right)^{-\gamma_{T_m}}\right). \tag{B.17}$$

In Table B.4, we report some Monte Carlo evidence on the size and power of our procedures using the approximations above. As can be seen, results are essentially the same as when using the Gumbel approximation; in this respect, this finding reinforces the statement in the paper by Gomes and Haan (1999) that when using the penultimate approximation, "the possible improvement is not spectacular".

Table B.4

Results with approximated critical values for the projected method, over 1000 replications, with significance level $\alpha = 0.05$, $k1 = k_2 = 3$, $m = p_2$, $r = 8$, $\epsilon = 0.05$, and $T = 200$. "PS", partial-sum; "WC", worst case.

			Empirical size ($\times 100$)		Empirical power ($\times 100$)		Empirical delay (interquantiles)	
T	p_1	p_2	PS ($\eta = 0.5$)	WC	PS ($\eta = 0.5$)	WC	PS ($\eta = 0.5$)	WC
200	20	80	2.9	5.6	76.9	73.9	16,21,28	14,18,25
200	30	80	2.6	6.9	97.7	96.3	11,14,20	9,12,17
200	40	80	2.6	6.7	99.8	99.8	9,11,14	7,10,12
200	80	20	2.6	5.0	24.5	39.8	6,8,10	5,6,8
200	80	30	3.3	6.8	94.5	96.4	6,8,10	5,6,8
200	80	40	2.9	6.2	99.8	100	5,7,9	4,6,7
200	20	20	8.8	7.6	62.8	67.6	5,7,10	4,5,7
200	30	30	3.9	7.4	94.3	95.8	5,7,10	4,5,7
200	40	40	3.0	6.7	99.7	99.9	5,6,9	4,5,7

B.1.6. *On the use of different sequences $\{z_\tau, 1 \leq \tau \leq T_m\}$.* In the construction of our CUSUM-based statistics $S_\tau = \sum_{j=0}^{\tau} y_j$, recall that we have employed $y_j = z_j + \psi_j$, with $z_j \sim i.i.d.N(0,1)$. Whilst this choice is quite natural, in this section we provide some alternative choices of z_j, also explaining in which way a Gaussian sequence may be more helpful. We begin with the following

Theorem A.3. *We assume that the assumptions of Theorems 1-2 hold, with $\{z_\tau, 1 \leq \tau \leq T_m\}$ i.i.d. with $E(z_\tau) = 0$, $E\left(z_\tau^2\right) = 1$ and $E|z_\tau|^\gamma < \infty$, for some $\gamma > 2$. Then, Theorems 1-2 hold.*

The theorem states that, asymptotically, any *i.i.d.* sequence with finite moments of order greater than 2 will yield the same result as Theorems 1-2. This result is a natural consequence of the KMT approximation for partial sums (Komlós et al., 1975, and Komlós et al., 1976); indeed, the proof of the theorem itself is a standard result which follows immediately from the KMT approximation, and more specifically on the following result

$$\max_{1\leq k\leq T_m}\frac{1}{k^{\zeta}}\left|\sum_{\tau=1}^{k}z_{\tau}-W_{T_m}\left(k\right)\right|=O_P\left(1\right),$$

which is required to hold for some $\zeta<1/2$. Such a result holds as long as $\gamma>2$, with

$$\zeta=1/2-1/\gamma,$$

which entails that (see also the passages in the proof) as γ increases, the approximation of the partial sums $S_{\tau}^{*}=\sum_{j=0}^{\tau}z_j$ with the Wiener process $W_{T_m}\left(\cdot\right)$ gets better. In fact, when the common distribution of the $\{z_{\tau},1\leq\tau\leq T_m\}$ has finite moment generating function in a finite neighborhood of zero, it holds that

$$\max_{1\leq k\leq T_m}\frac{\ln k}{k^{1/2}}\left|\sum_{\tau=1}^{k}z_{\tau}-W_{T_m}\left(k\right)\right|=O_P\left(1\right),$$

which is the best possible rate (Csörgo and Révész, 2014). Hence, using $z_j\sim i.i.d.N\left(0,1\right)$ (or, in general, z_j drawn from a distribution with finite moments of order higher than γ) yields the best approximation rates, and therefore $z_j\sim i.i.d.N\left(0,1\right)$ is a natural, "optimal" choice in this respect. Indeed, the same considerations would hold for the approximation of critical values for the case $\eta=1/2$ studied in Theorem A.2, and we refer to the comments at the end of the proof of that theorem for further discussion.

In Tables B.5, we report a small Monte Carlo exercise to show the impact of the distribution of z_j on the size and power of our procedure. As can be seen, results are virtually unchanged.

B.2. *Implementation guidelines.* In this section, we offer a further, more in-depth discussion of three issues which arise when implementing our methodologies: *(i)* how to estimate the number of common factors k_1 and k_2; *(ii)* the rationale for equation (4.1), which offers a selection rule for m; and *(iii)* how to choose r in the construction of (3.10). Our discussion complements the implementation guidelines in Section 4.1.

B.2.1. *Estimation of the number of common factors.* In real applications, k_1 and k_2 are usually unknown. Our results continue to hold as long as k_1 and k_2 are replaced by consistent estimators. From a practical point of view, we recommend using the estimators of k_1 and k_2 studied in He et al. (2023), which are shown to have very good finite sample properties.[3]

Corollary A.1. *Consider the estimator $\widetilde{k}_1$ and $\widetilde{k}_2$ such that $\widetilde{k}_1=k_1+o_{P^*}\left(1\right),$ and $\widetilde{k}_2=k_2+o_{P^*}(1)$. Then, Theorems 1-3 hold under the same assumptions when k_1 and k_2 are replaced by $\widetilde{k}_1$ and $\widetilde{k}_2$.*

[3]In particular, Theorem 4 in He et al. (2023) states that, as $\min\{p_1,p_2,m\}\to\infty$ the estimated numbers of factors, say $\widetilde{k}_1$ and $\widetilde{k}_2$, satisfy

$$\widetilde{k}_1=k_1+o_{P^*}\left(1\right),\quad\text{and}\quad\widetilde{k}_2=k_2+o_{P^*}\left(1\right),\tag{B.18}$$

for almost all realisations of $\{\mathbf{X}_t,1\leq t\leq T\}$.

Table B.5
Effect of the distribution of z_ψ for the projected method with "partial-sum", over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\epsilon = 0.05$, and $T = 200$.

			$z_\psi \overset{iid}{\sim} U[-\sqrt{3}, \sqrt{3}]$			$\frac{\sqrt{10}}{3} \times z_\psi \overset{iid}{\sim} t_{20}$		
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$
Empirical size ($\times 100$) under H_0								
200	40	80	4.2	1.1	2.2	4.4	1.9	2.8
200	80	40	4	1.6	2.9	3.3	1.4	2.5
200	40	40	4.8	1.0	2.5	5.2	1.9	3.2
Empirical power ($\times 100$) under H_{A1}								
200	40	80	99.7	99.8	99.6	99.7	99.7	99.7
200	80	40	99.9	99.8	99.9	99.9	99.9	99.7
200	40	40	99.7	99.6	99.0	99.8	99.7	99.1
Empirical delay (interquantiles) under H_{A1}								
200	40	80	9,11,14	8,11,14	10,12,15	9,11,14	8,11,14	9,12,15
200	80	40	6,7,9	6,7,9	7,8,11	5,7,9	5,7,9	7,8,10
200	40	40	5,6,9	5,7,9	6,8,10	5,7,9	5,7,9	6,8,10

The same result would hold also for more "classical" estimators $\widehat{k}_1$ and $\widehat{k}_2$, as long as $\widehat{k}_1 = k_1 + o_P(1)$ and $\widehat{k}_2 = k_2 + o_P(1)$. As a possible estimator, one could use the methodology developed in He et al. (2023).[4] Whilst the details are in that paper, here we report a brief discussion of how the estimators are constructed, in order to keep our exposition self-contained.

The estimators of (k_1, k_2), within the training period $1 \le t \le m$, are defined as $\left(\widetilde{k}_1, \widetilde{k}_2\right)$, and they are based on the projected matrix

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

defined in (2.1). Estimation is based on the j-th largest eigenvalue of $\widehat{\mathbf{M}}_1$, say $\widehat{\lambda}_j$, and on the transformation

$$\widehat{\phi}_j = \exp\left\{ \frac{p_1^{-\delta'} \widehat{\lambda}_j}{p_1^{-1} \sum_{j=1}^{p_1} \widehat{\lambda}_j} \right\} - 1, \tag{B.19}$$

where, similarly to (3.9)

$$\begin{cases} \delta' = \varepsilon' & \text{if } \beta \le 1/2 \\ \delta' = 1 - 1/(2\beta) + \varepsilon' & \text{if } \beta > 1/2 \end{cases}, \tag{B.20}$$

$\varepsilon' > 0$ is an arbitrarily small, user-defined number, and recall that $\beta = \ln p_1 / \ln(p_2 m)$. We then implement the following individual tests for

$$H_0 : k_1 \ge j \quad \text{vs.} \quad H_1 : k_1 < j$$

[4] We note that, in principle, also other estimators of (k_1, k_2) can be employed, and we mention, *inter alia*, the ones developed in Yu et al. (2022), Chen and Fan (2021), Han et al. (2022), and Chen and Lam (2022) - all these estimators are shown to work very well in the simulations in He et al. (2023).

based on a randomised test statistic constructed from $\widehat{\phi}_j$ according to the following algorithm.

Step 1 Generate *i.i.d.* samples $\{\eta^{(h)}\}_{h=1}^{M}$ with common distribution $N(0,1)$.

Step 2 Given $\{\eta^{(h)}\}_{h=1}^{M}$, construct sample sets $\{\widehat{\psi}_j^{(h)}(u)\}_{h=1}^{M}$ as

$$\widehat{\psi}_j^{(h)}(u) = I\left[\sqrt{\widehat{\phi}_j} \times \eta^{(h)} \leq u\right].$$

Step 3 Define

$$\widehat{\nu}_j(u) = \frac{2}{\sqrt{M}} \sum_{h=1}^{M} \left[\widehat{\psi}_j^{(h)}(u) - \frac{1}{2}\right].$$

Step 4 The test statistics are finally defined as

$$\widehat{\Psi}_j = \frac{1}{\sqrt{2\pi}} \int_U [\widehat{\nu}_j(u)]^2 \exp\left(-\frac{1}{2}u^2\right) du.$$

Hence, the estimator of k_1 (recall it is denoted as $\widetilde{k}_1$) is the output of the following algorithm:

Step 1 Run the test for $H_0 : k_1 \geq 1$ based on $\widehat{\Psi}_1$. If the null is rejected, set $\widetilde{k}_1 = 0$ and stop, otherwise go to the next step.

Step 2 For $j \geq 2$, run the test for $H_0 : k_1 \geq j$ based on $\widehat{\Psi}_j$, constructed using an artificial sample $\left\{\eta_j^{(h)}, 1 \leq h \leq M\right\}$ generated independently of $\left\{\eta_1^{(h)}, 1 \leq h \leq M\right\}$, ..., $\left\{\eta_{j-1}^{(h)}, 1 \leq h \leq M\right\}$. If the null is rejected, set $\widetilde{k}_1 = j-1$ and stop; otherwise repeat step 2 until the null is rejected, or until a pre-specified value $k_{\max}$ is reached.

The initial estimate, say $\widehat{k}_2$, is based on applying the same logic as above to the (eigenvalues of the) non-projected matrix $\mathbf{M}_c$ defined in (2.2). He et al. (2023) show that, as $\min\{p_1, p_2, m\} \to \infty$, it holds that $\widehat{k}_2 = k_2 + o_{P^*}(1)$ for almost all realisations of $\{\mathbf{X}_t, 1 \leq t \leq T\}$; thus, this estimator would suffice for our purposes. Indeed, upon obtaining $\widetilde{k}_1$, it is possible to iterate the procedure above using

$$\mathbf{M}_2 = \frac{1}{Tp_1} \sum_{t=1}^{T} \overline{\mathbf{Y}}_t' \overline{\mathbf{Y}}_t,$$

with $\overline{\mathbf{Y}}_t = \widetilde{\mathbf{R}}' \mathbf{X}_t / p_1$ and $\widetilde{\mathbf{R}}$ - as suggested by Yu et al. (2022) - can be set as $\widetilde{\mathbf{R}} = \sqrt{p_1}\mathbf{Q}^{\mathbf{R}}$, where the columns of $\mathbf{Q}^{\mathbf{R}}$ are the eigenvectors associated with the $\widetilde{k}_1$ largest eigenvalues of $\mathbf{M}_r$ defined in (2.2). The resulting estimator of k_2 is denoted as $\widetilde{k}_2$.

The simulations in He et al. (2023) indicate that $\left(\widetilde{k}_1, \widetilde{k}_2\right)$ delivers a superior performance compared to $\left(\widetilde{k}_1, \widehat{k}_2\right)$. Further, $\left(\widetilde{k}_1, \widetilde{k}_2\right)$ has the advantage that it uses the projection method, thus being consistent with the rest of the approach proposed in this paper, and also not requiring any further computations. Indeed, whilst the simulations in He et al. (2023) show that the finite sample performance of $\left(\widetilde{k}_1, \widetilde{k}_2\right)$ is, in general, very good, even better results for the estimation of the (crucial) quantity k_1 can be obtained by "deliberate overfitting" for k_2, i.e. by setting $\widetilde{k}_2 = k_{2,\max}$ with a pre-specified $k_{2,\max}$ (in He et al., 2023, the choice $k_{2,\max} = 8$ yields good results for all scenarios considered).

We now conclude this section with two sets of simulations. In Table B.6, we report results for the case where k_1 is not known; we use the true value of k_2, but unreported result show that this has little influence on the final outcomes (see also the simulations in Table B.7). As can be seen, results are essentially the same as if k_1 were known when p_1 and p_2 are large; results worsen when either p_1 (marginally) or p_2 (especially) is small, which was also noted in He et al. (2023), and it is due to the failure of the estimator of k_1 in those cases.

Finally, we also assess robustness to the estimation of k_2. In the main paper, recall that we have used the overfitting technique also mentioned above, with $\widetilde{k}_2 = k_{2,\max} = 8$. In Table B.7, we report results (for the method based on the projected estimator) obtained using $\widetilde{k}_2 = k_2$ (that is, assuming k_2 known) and using the estimated value $\widetilde{k}_2$. In both cases, results are comparable with the results obtained with $\widetilde{k}_2 = k_{2,\max} = 8$ reported in Tables 1-3 in the main paper, although, with small sample sizes, using $\widetilde{k}_2$ seems to yield better results.

B.2.2. *Further discussion on the choice of m.* In Section 4.1, we proposed a selection rule for m, designed so as to reduce the detection delay in the presence of a changepoint, viz.

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

presented in equation (4.1). Here, we report some algebra to explain the rationale of (4.1), and also discuss the case, which complements Assumption 3*(i)*, where $T_m = \Omega\left(m^{\varsigma}\right)$ with ς user-chosen and $\varsigma > 1$. We also present a comprehensive set of simulations to assess the impact of m on our methodology.

We begin by noting that Assumption 3*(ii)* requires

$$m\left|\left(\frac{p_1^{-\delta}}{p_2} + \frac{p_1^{-\delta}}{m} + \frac{p_1^{1-\delta}}{\sqrt{mp_2}}\right)\left(\ln^2 p_1 \ln p_2 \ln m\right)^{1+\epsilon}\right|^r \to 0.$$

Given that, by assumption, $r \geq 2$ and $\delta > 0$, it is easy to see that $m\left|\frac{p_1^{-\delta}}{m}\left(\ln^2 p_1 \ln p_2 \ln m\right)^{1+\epsilon}\right|^r \to 0$ always. We study

$$m\left|\frac{p_1^{1-\delta}}{\sqrt{mp_2}}\left(\ln^2 p_1 \ln p_2 \ln m\right)^{1+\epsilon}\right|^r \to 0.$$

This restriction is satisfied as long as

$$\left|\frac{p_1^{1-\delta}}{\sqrt{mp_2}}\right|^r = \Omega\left(\frac{1}{m^{1+\epsilon}}\right),$$

for all arbitrarily small $\epsilon > 0$. After some algebra, this is tantamount to requiring

$$\frac{p_1^{1-\delta}}{m} = \Omega\left(\frac{1}{m^{(1+\epsilon)/r}}\left(\frac{p_2}{m}\right)^{1/2}\right).$$

We know from (3.27) that the condition $\frac{p_1^{1-\delta}}{m} \to \infty$ ensures a short detection delay; this holds when

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

Table B.6

Effect of estimated k_1, over 1000 replications, with significance level $\alpha = 0.05$, $k1 = k_2 = 3$, k_2 given, $m = p_2$, $r = 8$, $\epsilon = 0.05$, and $T = 200$. "PS", partial-sum; "WC", worst case.

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical size ($\times 100$) under H_0										
200	20	80	10.3	7.4	8.8	11.0	40.4	27.6	13.5	13.7
200	30	80	4.2	1.4	2.3	4.2	40.0	24.9	8.9	6.9
200	40	80	4.4	1.9	2.7	5.1	46.0	24.8	9.9	7.9
200	80	20	12.4	9.1	10.4	10.8	12.0	8.8	10.5	10.9
200	80	30	4.2	1.4	3.3	4.0	4.2	2.1	3.8	3.7
200	80	40	3.4	1.7	3.4	5.2	3.9	1.1	2.9	4.2
200	20	20	37.3	33.9	33.0	31.9	75.4	65.7	48.2	43.3
200	30	30	4.3	1.9	2.5	4.4	63.4	48.9	23.1	13.2
200	40	40	3.8	1.5	2.5	3.4	60.4	45.4	18.4	9.2
Empirical power ($\times 100$) under H_{A1}										
200	20	80	81.0	77.6	72.6	74.2	90.4	86.6	73.5	68.1
200	30	80	98.0	97.1	95.9	96.0	98.7	98.6	95.1	92.5
200	40	80	100.0	99.9	99.8	99.7	100.0	99.8	99.7	99.6
200	80	20	34.0	28.7	22.8	42.4	14.3	11.3	10.6	17.2
200	80	30	95.0	94.7	90.8	96.1	65.9	57.9	45.9	68.4
200	80	40	100.0	99.8	99.6	99.8	96.3	95.6	93.1	95.6
200	20	20	76.3	71.9	63.9	76.9	91.6	85.7	70.8	77.0
200	30	30	95.3	94.1	91.6	95.2	98.4	97.5	91.4	92.8
200	40	40	99.9	99.8	99.6	100.0	100.0	99.8	99.0	98.9
Empirical delay (interquantiles) under H_{A1}										
200	20	80	15,20,29	15,20,28	15,21,28	12,17,24	16,22,32	15,22,31	16,23,32	14,19,26
200	30	80	11,14,20	11,14,19	12,16,21	10,13,17	13,16,21	12,16,22	14,18,23	11,15,20
200	40	80	9,11,14	9,11,14	10,12,16	8,10,13	10,13,16	10,12,16	11,14,18	9,11,15
200	80	20	3,8,11	-79,7,9	-74,6,10	3,6,8	-79,-77,14	-79,-79,-57	-74,-74,-74	-79,-25,8
200	80	30	6,8,10	6,8,10	8,9,12	5,6,8	9,12,15	9,11,14	10,12,15	7,9,11
200	80	40	5,7,9	6,7,9	7,8,10	4,6,7	8,10,13	8,10,13	10,12,15	7,8,11
200	20	20	-79,5,8	-79,5,8	-74,4,9	-79,3,6	-79,-18,8	-79,-37,7	-74,-72,8	-79,3,7
200	30	30	5,7,10	5,7,10	6,9,11	4,5,7	-6,6,9	-8,6,9	5,8,12	4,6,8
200	40	40	5,6,9	5,7,9	6,8,10	4,5,7	4,6,9	3,6,9	6,8,11	4,6,8

Table B.7

Effect of estimated k_2 for the projected method, over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = k_2 = 3$, k_1 given, $m = p_2$, $r = 8$, $\epsilon = 0.05$, and $T = 200$. "PS", partial-sum; "WC", worst case.

			k_2 is given				k_2 is estimated			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical size ($\times 100$) under H_0										
200	20	80	4.7	1.2	2.3	4.7	5.3	1.6	2.4	3.6
200	30	80	4.2	1.4	2.3	4.3	4.3	1.6	2.5	3.7
200	40	80	4.4	1.9	2.7	5.1	2.9	1.5	3.6	3.2
200	80	20	4.7	1.3	3.6	3.7	4.8	1.5	3.0	4.5
200	80	30	4.4	1.4	3.3	3.9	4.5	2.2	3.6	3.7
200	80	40	3.4	1.7	3.4	5.2	3.9	1.1	2.9	4.2
200	20	20	4.1	2.0	3.5	3.4	4.9	1.1	3.6	4.5
200	30	30	3.8	1.8	2.4	4.2	4.3	2.2	3.2	4.1
200	40	40	3.8	1.5	2.5	3.4	3.1	3.1	3.3	4.0
Empirical power ($\times 100$) under H_{A1}										
200	20	80	79.5	76.2	71.7	73.1	79.2	76.1	71.6	72.7
200	30	80	98.1	97.3	96.2	96.6	97.8	97.3	96.2	96.6
200	40	80	100.0	99.9	99.8	99.8	100.0	99.9	99.8	99.8
200	80	20	56.6	51.9	42.1	67.5	56.9	51.2	41.0	68.3
200	80	30	98.1	97.5	96.6	98.4	98.1	97.8	96.6	98.4
200	80	40	100.0	100.0	99.8	100.0	100.0	99.9	99.9	100.0
200	20	20	65.5	61.9	54.1	70.1	64.5	60.1	52.8	70.4
200	30	30	96.0	95.0	93.1	96.4	96.0	95.4	92.8	96.4
200	40	40	100.0	100.0	99.7	100.0	100.0	100.0	99.7	100.0
Empirical delay (interquantiles) under H_{A1}										
200	20	80	16,21,28	15,21,28	16,21,29	14,18,24	16,21,28	15,21,28	16,21,29	13,18,24
200	30	80	11,14,19	11,14,19	12,15,21	9,12,17	11,14,19	11,14,19	12,15,21	10,12,17
200	40	80	9,11,14	9,11,14	10,12,15	7,9,12	9,11,14	8,11,14	9,12,15	7,9,12
200	80	20	6,8,10	6,8,9	6,8,10	4,6,8	6,8,9	6,7,9	7,8,10	4,6,7
200	80	30	5,7,9	5,7,9	7,8,10	4,5,7	5,7,9	5,7,9	7,8,10	4,5,7
200	80	40	5,6,8	5,6,8	6,8,9	4,5,7	5,6,8	5,6,8	6,8,9	4,5,7
200	20	20	5,7,9	5,7,9	6,8,10	4,5,7	5,7,9	5,7,9	6,8,10	4,5,7
200	30	30	5,7,9	5,7,9	6,8,10	4,5,7	5,7,9	5,7,9	6,8,10	4,5,7
200	40	40	5,6,8	5,6,8	6,7,9	4,5,7	5,6,8	5,6,8	6,7,10	4,5,7

Since the above entails that $m = o(p_2^r)$, it also readily follows that

$$m \left| \frac{p_1^{-\delta}}{p_2} \left(\ln^2 p_1 \ln p_2 \ln m\right)^{1+\epsilon} \right|^r \to 0,$$

thus ensuring the validity of (3.12).

Similar considerations would allow to study the case where, in Assumption 3*(i)*, we require $T_m = \Omega(m^\varsigma)$, $\varsigma \geq 1$. In such a case, Assumption 3*(ii)* requires

$$m^\varsigma \left| \left(\frac{p_1^{-\delta}}{p_2} + \frac{p_1^{-\delta}}{m} + \frac{p_1^{1-\delta}}{\sqrt{mp_2}} \right) \left(\ln^2 p_1 \ln p_2 \ln m\right)^{1+\epsilon} \right|^r \to 0.$$

In order for this to hold, it is required that

$$\varsigma < r. \tag{B.21}$$

Moreover, from studying

$$m^\varsigma \left| \frac{p_1^{1-\delta}}{\sqrt{mp_2}} \left(\ln^2 p_1 \ln p_2 \ln m\right)^{1+\epsilon} \right|^r \to 0,$$

the same passages above suggest the selection rule

$$m = O\left(p_2^{r/(r+2\varsigma)-\epsilon}\right), \tag{B.22}$$

for all arbitrarily small $\epsilon > 0$. It is now easy to see that

$$m^\varsigma \left| \frac{p_1^{-\delta}}{p_2} \left(\ln^2 p_1 \ln p_2 \ln m\right)^{1+\epsilon} \right|^r \to 0.$$

As mentioned in Section 4.1, the "optimal" rule $m = O\left(p_2^{r/(r+2)-\epsilon}\right)$ may, in practice, yield a small m, which in turn may not afford a good asymptotic approximation under the null. Hence, in practice, especially if p_2 is small, a sub-optimal value of m may be preferable. According to (3.27) it is required that

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

if m is "too large", i.e. if (4.1) does not hold, the condition above can still be satisfied, but the delay $\tau - t^*$ may also need to grow with m before detection takes place. In Section B.3.3, we report a small Monte Carlo exercise to show the impact of m on the size and power of our procedure.

B.2.3. *Further discussion on the choice of r.* In Section 3, we have constructed ψ_τ using

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

As we pointed out, under the null, we need ψ_τ to drift to zero as fast as possible, and this requires a large value of r; under the alternative, we need ψ_τ to diverge to positive infinity as fast as possible, and again this requires a large value of r. The lack of a trade-off between the behaviour under the null and under the alternative, therefore, suggests that r should be as large as possible.

In Section 4.1, we have discussed a possible approach to choose r, based on the data $\mathbf{X}_t$. In particular, if

$$E\left|X_{ij,t}\right|^{4\widetilde{r}} < \infty, \tag{B.23}$$

for all $1 \leq i \leq p_1$, $1 \leq j \leq p_2$, and $1 \leq t \leq m$, then this suggests the choice $r = \widetilde{r}$. We point out that, in principle, it is possible to use the test by Trapani (2016) for the following hypothesis testing problem:

$$\left\{ \begin{array}{ll} H_0: & E\left|X_{ij,t}\right|^{4\widetilde{r}} = \infty \\ H_A: & E\left|X_{ij,t}\right|^{4\widetilde{r}} < \infty \end{array} \right. . \tag{B.24}$$

As suggested in Trapani (2016), the test can be based on the (rescaled) sample moment

$$\widehat{\mu}_{p_1,p_2,m}^{\widetilde{r}} = \frac{\frac{1}{mp_1p_2}\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\sum_{t=1}^{m}\left|X_{ij,t}\right|^{4\widetilde{r}}}{\left(\frac{1}{mp_1p_2}\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\sum_{t=1}^{m}X_{ij,t}^2\right)^{2\widetilde{r}}}.$$

For the sake of a self-contained exposition, we report the (high-level) assumptions that are required for the test by Trapani (2016) to be applicable.

Assumption A.6. *Let $\gamma^{\max}$ be the largest number such that*

$$E\left|X_{ij,t}\right|^{\gamma^{\max}} < \infty,$$

for all $1 \leq i \leq p_1$, $1 \leq j \leq p_2$, and $1 \leq t \leq m$. Then it holds that (i)

$$\frac{1}{mp_1p_2}\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\sum_{t=1}^{m}\left|X_{ij,t}\right|^{\gamma} \to c_0 \in (0,\infty),$$

for all $\gamma \leq \gamma^{\max}$; (ii) there exists a constant $0 < c_0 < \infty$, and a triplet of random variables $\{m_0, p_{1,0}, p_{2,0}\}$ such that, for all $m \geq m_0$, $p_1 \geq p_{1,0}$ and $p_2 \geq p_{2,0}$

$$\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\sum_{t=1}^{m}\left|X_{ij,t}\right|^{\gamma} \geq c_0\frac{\left(mp_1p_2\right)^{\gamma/\gamma^{\max}}}{v\left(m,p_1,p_2\right)},$$

for all $\gamma > \gamma^{\max}$, where $v\left(m,p_1,p_2\right)$ is a slowly varying sequence.

As mentioned above, Assumption A.6 could be verified e.g. upon assuming a specific form of serial and spatial dependence for $\{X_{ij,t}\}$. In particular, part *(i)* is a Strong Law of Large Numbers, which could be shown from routine arguments assuming weak dependence across i, j and t. As far as part *(ii)* is concerned, this provides an a.s. lower bound to the divergence of $\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\sum_{t=1}^{m}\left|X_{ij,t}\right|^{\gamma}$. Again, this could be in principle derived from more primitive conditions: for example, Barigozzi et al. (2022) show a similar result (see their Proposition 2) for a single time series, assuming strong mixing, and the main tool required in their proof is an inequality for the moments of partial sums of strongly mixing variables. If such an inequality could hold for the random field $\{X_{ij,t}, 1 \leq i \leq p_1, 1 \leq j \leq p_2, 1 \leq t \leq m\}$, then the proof of Proposition 2 in

Barigozzi et al. (2022) could be repeated *verbatim*, yielding part *(ii)* of Assumption A.6 with $v(m, p_1, p_2) = (\ln(mp_1p_2))^{(2+\epsilon)\gamma/\gamma^{\max}}$, valid for all $\epsilon > 0$. In principle, therefore, part *(ii)* could be shown upon assuming some specific form of dependence for $\{X_{ij,t}\}$; for example, Miller (1994) and Zhang (1998) study the growth rates of partial sums for uniform and ρ-mixing random fields. One difficulty with this approach, though, is that it requires a specific assumption on the form of dependence: whilst such assumption may be plausible as far as dependence across time is concerned, the absence of a natural ordering across i and j makes it more difficult to justify a specific form of dependence across those two indices, albeit in principle possible.

Based on the rates entailed by Assumption A.6*(i)* under the alternative hypothesis H_A in (B.24), and Assumption A.6*(i)* under the null hypothesis H_0, the test in Trapani (2016) can be readily applied to test for (B.24).

In Table B.8, we report a small Monte Carlo exercise to show the impact of r in (3.10) on the size and power of our procedure. As can be seen, results are robust to a wide variety of values of r when using the projected estimator.

B.3. *Further Monte Carlo evidence.* In this section, we consider the sensitivity of our results to several specifications of our procedure, and we complement the theory developed in Sections B.1 and B.2.

B.3.1. *Sensitivity to k_2.* In the simulations in the main paper, we have used $k_2 = 3$. In this subsection, we consider the impact of different values of k_2, trying $k_2 = 2$ and $k_2 = 4$. All the other settings are the same as those in the main paper.

The empirical sizes, and the empirical powers and delays under the alternative $H_{A,1}$ in (3.2) are reported in Table B.9 - we have also carried out some experiments under $H_{A,2}$, but results are in line with the ones under $H_{A,1}$ and the results in Section 4.2. The results broadly show that changing k_2 has virtually no impact on our methodology when using the projection-based estimator; conversely, results are more sensitive when using the estimation method without projection. This is more likely to lose control of size when k_2 is smaller, whereas when k_2 is larger, the size is better controlled.

B.3.2. *Sensitivity to δ.* In Section 4.1, we recommend setting $\varepsilon = 0.05$ in the construction of δ defined in (3.9). Here, we consider the sensitivity of our results to ε, using $\varepsilon \in \{0.02, 0.03, 0.04, 0.06\}$. All the other settings are the same as those in the main paper.

The results in Table B.10 show that varying ε has virtually no impact.

B.3.3. *Sensitivity to m.* In Section 4.1, we offer an upper bound for m which should deliver a short detection delay; however, when p_2 is small, such an upper bound may not be the best choice in practice. Moreover, the upper bound in (4.1) does not offer an adaptive rule to select m *per se*, but merely a bound. Here, we investigate how our procedure is affected by varying m.

The results in Table B.11 show that, when using the projection-based estimator, the empirical size is virtually unaffected by m, and that - although the power decreases and the detection delay increases as m grows, as expected - the impact of m is still only marginal. Conversely, when using the estimation method without projection, this yields either a loss of size control, or lower power/longer delay.

B.3.4. *Sensitivity to the size of change.* We complement our results in the main paper by assessing how the power of our procedure changes as the magnitude of the changepoint changes - results

Table B.8

Effect of transformation function $g(x) = |x|^r$, over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = 3$ (given), $m = p_2$, $r = 8$, $\epsilon = 0.05$, and $T = 200$. "PS", partial-sum; "WC", worst case.

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
p_1	p_2	r	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical size (×100) under H_0										
80	40	2	7.2	2.0	2.7	5.2	48.7	23.7	6.4	7.4
80	40	4	4.9	2.0	2.6	3.7	4.5	2.4	3.1	5.5
80	40	6	4.1	1.3	2.8	4.8	3.4	1.6	2.4	3.2
80	40	8	4.7	1.8	1.4	3.4	4.1	1.6	3.3	3.6
80	40	10	4.1	1.6	2.4	3.9	5.1	2.1	3.2	4.2
80	40	12	4.4	2.2	3.2	5.1	4.0	1.6	1.7	4.2
40	40	2	42.7	21.1	6.1	7.3	100.0	100.0	98.6	34.0
40	40	4	4.7	2.9	3.3	5.5	98.3	93.8	62.0	21.0
40	40	6	4.3	1.9	2.9	4.0	83.6	69.1	31.6	14.7
40	40	8	3.6	2.1	4.1	4.8	61.6	44.1	18.9	10.5
40	40	10	4.7	1.8	2.5	4.4	41.9	29.9	12.4	8.1
40	40	12	4.2	1.1	4.0	3.7	29.0	19.5	8.9	8.2
Empirical power (×100) under H_{A1}										
80	40	2	100.0	99.7	96.5	98.0	99.4	98.3	82.3	78.7
80	40	4	100.0	99.6	99.1	99.6	96.8	94.7	84.1	91.4
80	40	6	99.8	99.6	99.6	99.8	95.8	94.8	89.8	94.8
80	40	8	99.9	99.9	99.7	99.9	97.0	96.0	93.0	96.4
80	40	10	99.9	99.9	99.8	99.8	97.0	95.7	94.3	97.2
80	40	12	100.0	100.0	100.0	100.0	97.8	97.4	95.1	98.0
40	40	2	99.9	99.7	98.1	97.6	100.0	100.0	100.0	92.5
40	40	4	99.6	99.2	98.1	99.1	100.0	99.9	99.8	97.0
40	40	6	99.8	99.7	99.2	99.7	100.0	100.0	99.2	98.9
40	40	8	99.7	99.6	99.4	99.5	99.8	99.8	99.2	99.0
40	40	10	99.9	99.9	99.9	99.8	99.8	99.8	99.1	99.6
40	40	12	99.9	99.7	99.5	99.9	99.7	99.7	99.3	99.6
Empirical delay (interquantiles) under H_{A1}										
80	40	2	9,12,14	10,12,15	13,16,19	7,9,11	9,13,16	11,14,18	15,19,23	9,12,16
80	40	4	7,9,11	7,9,12	9,11,14	5,7,9	10,13,16	11,13,17	13,15,19	7,10,13
80	40	6	6,8,10	6,8,10	8,9,11	5,6,8	9,11,14	9,12,15	11,13,17	7,9,11
80	40	8	5,7,9	5,7,9	7,8,10	4,6,7	8,10,13	8,11,13	9,12,15	6,8,10
80	40	10	5,7,8	5,7,8	6,8,10	4,6,7	8,10,12	8,10,12	9,11,14	6,8,10
80	40	12	5,6,8	5,6,8	6,8,9	4,5,7	7,9,12	7,9,12	8,11,13	6,8,10
40	40	2	7,9,12	7,10,13	11,13,17	6,8,11	-38,-30,-21	-48,-39,-26	-52,-40,-13	6,9,13
40	40	4	6,8,11	6,8,11	8,10,13	4,6,8	-28,-11,5	-37,-14,6	-42,7,11	5,7,10
40	40	6	5,7,9	5,7,9	7,9,11	4,5,7	-10,5,8	-15,6,9	5,8,11	5,6,9
40	40	8	5,7,8	5,7,9	6,8,10	4,5,7	4,6,9	4,6,9	6,8,11	4,6,8
40	40	10	5,6,8	5,6,8	6,7,10	4,5,7	4,7,9	4,7,9	6,8,11	4,6,8
40	40	12	4,6,8	4,6,8	6,7,9	4,5,7	5,7,9	5,7,9	6,8,11	4,6,8

TABLE B.9

Effect of k_2, over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = 3$, $m = p_2$, $r = 8$, $\epsilon = 0.05$, and $T = 200$. "PS", partial-sum; "WC", worst case.

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
k_2	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical size ($\times 100$) under H_0										
2	40	80	4.1	2.3	3.3	5.0	98.2	94.2	80.4	42.1
2	80	40	4.4	1.7	2.1	4.0	3.1	1.3	3.1	3.7
2	40	40	5.2	1.9	2.7	4.4	99.1	98.5	90.8	61.7
4	40	80	4.1	1.8	3.2	5.1	7.7	2.9	3.3	4.7
4	80	40	4.6	1.3	3.1	3.9	4.6	1.7	2.8	4.7
4	40	40	5.0	2.2	2.8	3.7	14.4	6.4	4.1	5.3
Empirical power ($\times 100$) under H_{A1}										
2	40	80	100.0	100.0	99.9	100.0	100.0	100.0	99.9	99.7
2	80	40	99.7	99.6	99.4	99.7	89.3	86.3	78.8	86.6
2	40	40	99.7	99.6	98.9	99.4	100.0	100.0	99.9	99.3
4	40	80	99.9	99.9	99.9	99.8	99.8	99.6	99.6	99.5
4	80	40	100.0	100.0	99.9	100.0	98.4	98.1	97.0	98.0
4	40	40	100.0	99.8	99.7	99.7	100.0	99.9	98.8	99.5
Empirical delay (interquantiles) under H_{A1}										
2	40	80	9,11,15	8,11,15	9,13,16	7,10,13	-4,5,11	-11,-1,10	-14,-1,13	7,11,16
2	80	40	5,7,9	6,7,10	7,9,11	4,6,8	10,12,16	10,12,16	11,14,18	7,10,13
2	40	40	5,7,9	5,7,10	6,9,11	4,6,8	-47,-37,-21	-54,-45,-26	-54,-49,-13	-31,4,8
4	40	80	9,11,14	8,11,14	9,12,15	7,9,12	10,13,16	10,12,16	11,14,18	9,11,14
4	80	40	5,7,8	5,7,8	7,8,10	4,5,7	7,9,11	7,9,12	9,11,13	6,7,9
4	40	40	5,6,8	5,7,9	6,8,10	4,5,7	6,7,10	6,8,10	7,9,11	4,6,8

Table B.10

Effect of ε, over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, and $T = 200$. "PS", partial-sum; "WC", worst case.

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
ϵ	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical size ($\times 100$) under H_0										
0.02	80	40	4.9	1.9	4.3	3.8	3.2	1.2	2.5	4.4
0.04	80	40	4.8	2.0	3.0	3.6	4.1	1.9	3.2	4.0
0.06	80	40	4.1	1.6	3.2	3.7	4.6	1.3	2.7	3.5
0.08	80	40	4.7	2.0	3.7	4.6	4.1	1.8	3.5	4.6
0.02	40	40	6.2	2.7	3.8	4.4	94.5	88.9	65.0	32.2
0.04	40	40	4.7	2.2	3.2	4.1	76.8	60.9	32.3	14.8
0.06	40	40	3.5	2.1	3.4	4.0	44.3	28.2	10.2	8.3
0.08	40	40	3.8	2.1	3.3	4.3	23.1	10.9	5.3	6.1
Empirical power ($\times 100$) under H_{A1}										
0.02	80	40	100.0	100.0	100.0	100.0	99.0	98.8	98.2	98.8
0.04	80	40	100.0	100.0	100.0	100.0	97.4	96.4	94.6	97.1
0.06	80	40	99.9	99.9	99.8	99.9	95.1	93.9	89.5	94.0
0.08	80	40	99.3	99.1	98.5	99.0	87.8	84.9	77.7	86.2
0.02	40	40	99.9	99.9	99.8	99.9	100.0	100.0	100.0	100.0
0.04	40	40	99.8	99.8	99.8	99.8	99.9	99.9	99.6	99.7
0.06	40	40	99.8	99.6	99.2	99.7	99.5	99.2	98.1	98.6
0.08	40	40	99.2	99.0	98.0	99.1	98.5	97.1	94.0	95.8
Empirical delay (interquantiles) under H_{A1}										
0.02	80	40	5,6,8	5,6,8	6,7,9	4,5,6	7,9,11	7,9,11	8,10,13	5,7,9
0.04	80	40	5,7,8	5,7,9	7,8,10	4,5,7	8,10,12	8,10,13	9,11,14	6,8,10
0.06	80	40	6,7,9	6,7,9	7,9,11	4,6,7	9,11,14	9,11,14	10,12,15	7,9,11
0.08	80	40	6,8,10	6,8,10	8,10,12	5,7,8	10,12,15	10,12,15	11,13,17	8,10,12
0.02	40	40	4,6,8	4,6,8	5,7,9	3,5,6	-34,-16,4	-44,-20,4	-51,-5,7	3,4,6
0.04	40	40	5,6,8	5,6,8	6,8,10	4,5,7	-12,5,7	-18,5,8	4,7,10	4,6,8
0.06	40	40	5,7,9	5,7,9	6,8,11	4,5,7	5,7,10	5,7,10	7,9,12	5,6,9
0.08	40	40	6,8,10	6,8,10	7,9,12	5,6,8	7,9,12	7,9,12	9,11,14	6,7,10

Table B.11

Effect of m, over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = k_2 = 3$, $r = 8$, $\epsilon = 0.05$ and $T = 200$. "PS", partial-sum; "WC", worst case.

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
m	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical size ($\times 100$) under H_0										
20	80	40	5.2	1.9	4.3	3.7	2.9	1.4	2.2	4.2
30	80	40	5.0	1.9	3.0	3.4	4.0	1.7	4.0	4.0
40	80	40	4.1	1.6	3.2	3.7	4.6	1.2	2.7	3.5
50	80	40	4.9	2.0	3.7	4.9	3.9	1.3	2.9	4.8
20	40	40	6.1	2.4	3.9	4.4	5.3	1.7	2.6	5.0
30	40	40	4.8	2.1	3.2	3.9	18.2	7.5	4.3	5.1
40	40	40	3.5	2.1	3.4	4.0	58.4	42.3	18.8	10.9
50	40	40	3.4	2.1	3.3	4.5	57.9	39.4	17.7	10.5
Empirical power ($\times 100$) under H_{A1}										
20	80	40	89.6	87.3	81.5	93.5	40.5	34.0	25.9	51.7
30	80	40	99.0	98.6	97.1	99.2	82.0	78.8	71.6	84.2
40	80	40	99.9	99.9	99.9	99.9	96.8	95.6	93.0	95.8
50	80	40	100.0	100.0	100.0	100.0	99.2	99.0	98.2	98.6
20	40	40	88.3	85.7	81.2	91.8	67.7	60.6	50.7	71.9
30	40	40	98.7	98.3	97.1	99.1	97.0	95.8	90.7	95.1
40	40	40	99.8	99.8	99.5	99.8	100.0	99.7	99.2	99.2
50	40	40	100.0	99.9	99.4	99.8	99.7	99.7	99.1	99.0
Empirical delay (interquantiles) under H_{A1}										
20	80	40	4,6,8	5,6,8	6,7,9	3,5,6	7,9,11	7,8,11	7,9,11	5,6,8
30	80	40	5,7,8	5,7,8	6,8,10	4,5,7	8,10,12	8,10,12	9,11,14	6,7,9
40	80	40	6,7,9	6,7,9	7,8,10	4,6,7	8,10,13	8,10,13	10,12,15	6,8,10
50	80	40	6,8,9	6,8,9	7,9,11	5,6,8	9,11,13	9,11,14	10,12,15	7,9,11
20	40	40	4,5,7	4,6,7	5,7,9	3,4,6	5,7,10	5,7,9	6,8,10	4,5,7
30	40	40	5,6,8	5,6,8	6,8,10	4,5,6	6,7,10	6,8,10	7,9,12	4,6,8
40	40	40	5,7,9	5,7,9	6,8,10	4,5,7	3,6,9	4,6,9	6,8,11	4,6,8
50	40	40	6,8,10	6,8,11	7,9,12	5,7,9	5,8,11	5,8,11	8,10,13	6,8,10

are in Table B.12. Recall that, when considering $H_{A,1}$, after the changepoint we have re-generated a fraction γ of entries of the k_1-th column of $\mathbf{R}$ from a uniform distribution $\mathcal{U}(-1, 1)$. Similarly, under $H_{A,2}$, after t^* we added a new column to $\mathbf{R}$, with a fraction γ of its entries sampled from a uniform distribution $\mathcal{U}(-1, 1)$, and the other entries set equal to zero. Thus, the size of the break is controlled by γ, increasing as γ increases.

B.3.5. *Empirical size and power in the presence of changes in* $\mathbf{C}$. We investigate the effects on our monitoring procedure of changes in the space spanned by the column factor loadings $\mathbf{C}$, complementing the theory studied in Section B.1.4.

We begin by investigating the empirical rejection frequencies under the null of no change in $\mathbf{R}$, generating data according to

$$\mathbf{X}_t = \begin{cases} \mathbf{R}\mathbf{F}_t\mathbf{C}_1' + \mathbf{E}_t, & 1 \le t \le t^*, \\ \mathbf{R}\mathbf{F}_t\mathbf{C}_2' + \mathbf{E}_t, & t^*+1 \le t \le T, \end{cases} \tag{B.25}$$

where $\mathbf{C}_1$ is generated with *i.i.d.* entries sampled from $\mathcal{U}(-1, 1)$, and $\mathbf{C}_2$ is generated after a time point t^*+1 with *i.i.d.* entries from $\mathcal{U}(-1, 1)$; the time point t^* is chosen as in the other experiments as $t^* = 0.5T$. All the other parameters are set to the same values as those introduced in Section 4.2. The empirical sizes over $1,000$ replications are reported in Table B.13, and, comparing these results with Table 1 in the main paper, we conclude that the monitoring procedures have controlled empirical sizes no matter whether the column loading matrix changes or not, i.e, the change of the loading matrix $\mathbf{C}$ have limited impact on the empirical sizes when we detect changes for loading matrix $\mathbf{R}$.

We have also considered power under the alternative (3.2) - i.e. the case where the number of common factors increases after the change point.[5] Data are generated similarly to Section 4.2, using

$$\mathbf{X}_t = \begin{cases} \mathbf{R}_1\mathbf{F}_{1,t}\mathbf{C}_1' + \mathbf{E}_t, & 1 \le t \le t^*, \\ \mathbf{R}_2\mathbf{F}_{1,t}\mathbf{C}_2' + \mathbf{E}_t, & t^*+1 \le t \le T, \end{cases}, \tag{B.26}$$

where $\mathbf{R}_2$ is generated after time point t^*+1, with a fraction γ of its entries sampled from *i.i.d.* uniform distribution $\mathcal{U}(-1, 1)$ (the remaining entries are the same as in $\mathbf{R}_1$), and $\mathbf{C}_2$ is also generated from the time point t^*+1 with *i.i.d.* entries from $\mathcal{U}(-1, 1)$. We point out that, in (B.26), the row and column loading spaces change at the same time t^*+1 merely for simplicity - in unreported experiments, we considered different times of change for $\mathbf{R}$ and $\mathbf{C}$, noting that this has little effects on the results. As above, all the other parameters are set to the same values as those introduced in Section 4.2. Compared with Tables 2 and 3 in the main paper, the results in the two bottom panels of Table B.13 show that the presence of changes in $\mathbf{C}$ has limited effects on the procedures.

B.3.6. *Power versus smooth transitioning breaks.* In the main paper, we have studied the performance of our methodologies versus abrupt breaks. However, the case of slowly transitioning changes in the loadings matrices is also of great interest and empirical relevance.[6] Bates et al. (2013) analyse the case of time-varying loadings in the context of a vector factor model.

In this section, we study the empirical rejection frequencies under two cases. We consider the case - denoted as $H_{A,3}$ - in which

$$\mathbf{X}_t = \begin{cases} \mathbf{R}_0\mathbf{F}_t\mathbf{C}' + \mathbf{E}_t, & 1 \le t \le m+k^* \\ \mathbf{R}_t\mathbf{F}_t\mathbf{C}' + \mathbf{E}_t, & m+k^*+1 \le t \le T \end{cases}, \tag{B.27}$$

[5] We have also considered the case of alternative (3.4) in unreported experiments, obtaining the same results.

[6] We are grateful to an anonymous Referee for pointing this out to us.

Table B.12

Effect of γ, over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\epsilon = 0.05$ and $T = 200$. "PS", partial-sum; "WC", worst case.

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
γ	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical power ($\times 100$) under H_{A1}										
0.10	80	20	4.2	2.8	2.7	3.4	3.9	1.4	2.3	3.9
0.15	80	20	4.7	1.8	2.3	6.0	4.4	1.6	3.8	4.2
0.20	80	20	9.4	5.1	5.2	10.3	4.5	2.2	3.4	4.3
0.25	80	20	15.6	11.1	9.4	18.5	6.8	2.3	2.3	6.4
0.30	80	20	35.3	29.5	19.1	45.1	10.0	5.6	3.0	12.2
0.10	80	30	11.0	6.6	6.4	11.0	5.1	1.8	3.1	4.1
0.15	80	30	38.4	33.1	26.7	38.9	11.3	7.0	5.8	11.4
0.20	80	30	61.1	55.7	46.7	63.1	20.5	15.5	10.9	20.7
0.25	80	30	85.9	82.7	73.9	86.7	42.2	33.4	25.0	42.3
0.30	80	30	96.5	95.3	92.8	96.3	68.7	61.8	51.5	72.4
0.10	80	40	34.0	29.3	24.4	33.4	10.8	6.4	5.9	8.5
0.15	80	40	83.4	79.5	74.3	81.7	46.7	38.3	29.2	41.3
0.20	80	40	94.0	91.6	89.4	93.5	68.6	62.5	52.0	65.9
0.25	80	40	99.3	99.2	98.2	99.4	89.2	86.0	78.0	85.9
0.30	80	40	100.0	100.0	100.0	100.0	98.1	97.7	95.9	97.5
Empirical delay (interquantiles) under H_{A1}										
0.10	80	20	-3,32,68	-78,-58,-37	-74,-74,-72	-4,16,70	40,59,80	-77,-76,-58	-74,-74,-74	-22,28,67
0.15	80	20	10,23,64	-62,-6,12	-74,-74,-72	-10,8,45	17,45,80	-60,-50,-34	-74,-74,-72	-51,-10,51
0.20	80	20	8,12,48	-26,8,12	-74,-73,-29	5,7,10	1,31,67	-75,-66,-38	-74,-74,-73	-34,11,50
0.25	80	20	8,10,17	7,9,12	-72,8,10	5,7,8	11,31,51	-15,9,24	-74,-73,-72	6,8,24
0.30	80	20	7,9,11	7,9,11	7,9,11	5,6,8	8,12,52	-12,8,13	-74,-71,10	5,7,9
0.10	80	30	12,18,47	3,12,18	-64,-62,14	8,13,18	12,32,68	-62,-22,10	-64,-64,-63	-20,23,64
0.15	80	30	10,13,16	9,13,15	10,13,16	7,10,12	12,18,42	-13,12,17	-64,-56,15	7,11,16
0.20	80	30	9,12,15	9,11,15	10,12,16	7,9,12	12,15,21	11,13,17	7,14,17	8,10,13
0.25	80	30	7,10,13	7,10,13	9,11,14	6,8,10	10,13,18	10,12,15	10,13,15	7,10,12
0.30	80	30	6,8,10	6,8,10	7,9,11	4,6,8	8,11,14	8,10,13	9,12,14	6,8,11
0.10	80	40	14,17,22	13,16,20	13,17,21	11,14,17	19,26,59	10,18,24	-54,-47,20	11,16,22
0.15	80	40	10,13,17	10,13,17	11,14,18	8,10,13	14,18,23	14,17,21	13,18,21	11,14,17
0.20	80	40	8,11,14	8,10,14	9,12,15	6,9,11	11,15,19	11,15,18	12,15,20	9,12,15
0.25	80	40	6,8,11	6,9,11	8,10,13	5,7,9	10,12,16	10,12,16	11,14,17	7,10,13
0.30	80	40	5,7,8	5,7,8	7,8,10	4,5,7	8,10,12	8,10,13	9,11,14	6,8,10

Table B.13
Results when C also changes, over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\epsilon = 0.05$ and $T = 200$. "PS", partial-sum; "WC", worst case.

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical size ($\times 100$) under H_0										
200	20	80	4.1	2.3	3.1	3.4	36.0	21.9	7.6	8.2
200	30	80	4.5	2.2	3.7	5.0	39.5	23.5	7.1	7.0
200	40	80	4.0	2.1	3.0	4.2	43.9	26.5	8.1	7.7
200	80	20	3.2	1.3	2.4	4.9	5.2	1.2	2.7	3.2
200	80	30	3.8	1.7	2.6	4.6	3.5	1.8	3.5	3.8
200	80	40	4.0	1.7	3.0	3.8	3.6	1.8	3.2	4.3
200	20	20	9.7	5.2	4.6	5.0	64.1	52.0	31.0	19.7
200	30	30	4.7	1.9	2.9	3.2	66.4	49.0	23.7	15.0
200	40	40	3.7	1.9	3.4	3.9	63.9	44.5	15.7	10.0
Empirical power ($\times 100$) under H_{A1}										
200	20	80	81.7	79.1	74.3	75.6	91.5	86.4	72.7	69.4
200	30	80	98.1	97.3	96.2	96.3	98.9	97.8	94.5	92.5
200	40	80	99.9	99.9	99.5	99.6	99.8	99.7	99.5	98.8
200	80	20	30.0	22.5	17.3	39.6	7.2	3.9	3.4	9.0
200	80	30	95.4	93.4	89.6	95.6	64.1	57.8	46.9	65.9
200	80	40	99.9	99.9	99.6	99.9	97.1	95.8	92.8	96.5
200	20	20	63.1	55.8	44.5	63.7	89.0	81.4	58.7	62.6
200	30	30	94.9	93.6	90.8	95.2	98.4	96.9	91.1	92.3
200	40	40	99.6	99.5	99.0	99.8	99.7	99.6	98.9	98.9
Empirical delay (interquantiles) under H_{A1}										
200	20	80	16,21,29	16,21,28	17,22,29	14,19,24	17,23,33	17,23,32	18,24,33	16,21,29
200	30	80	11,15,19	11,15,19	12,16,21	10,13,17	12,16,22	12,16,22	14,18,24	11,15,20
200	40	80	8,11,14	8,11,14	9,12,15	7,10,12	9,13,16	9,12,16	10,14,18	8,11,15
200	80	20	7,10,14	7,9,11	6,9,11	5,7,9	10,20,57	-56,7,14	-74,-73,-69	5,8,14
200	80	30	6,8,10	6,8,10	7,9,12	5,6,8	9,11,15	9,11,14	9,12,15	6,9,11
200	80	40	5,7,9	5,7,9	7,8,10	4,6,7	8,10,13	8,10,13	10,12,15	6,8,11
200	20	20	5,7,10	5,7,10	6,8,10	4,5,7	-28,5,11	-38,5,11	-56,7,11	3,5,8
200	30	30	5,7,9	5,7,9	6,8,11	4,5,7	-8,5,9	-10,6,9	5,8,11	4,6,8
200	40	40	5,6,9	5,7,9	6,8,10	4,5,7	4,6,9	4,7,9	6,9,11	4,6,8

where

$$\mathbf{R}_t = \mathbf{R}_0 + \theta \widetilde{\mathbf{R}}_t. \tag{B.28}$$

This case should not be viewed as a break, because it corresponds to a matrix factor model with no changes in the row column loadings space, and heteroskedastic factors. Similar considerations would apply to a model with a smooth change with non-random design, viz.

$$\mathbf{X}_t = \begin{cases} \mathbf{R}_0 \mathbf{F}_t \mathbf{C}' + \mathbf{E}_t, & 1 \leq t \leq m, \\ \mathbf{R}_t \mathbf{F}_t \mathbf{C}' + \mathbf{E}_t, & m+1 \leq t \leq T, \end{cases} \tag{B.29}$$

where

$$\mathbf{R}_t = \mathbf{R}_0 \left(1 + \frac{t}{m} \right). \tag{B.30}$$

Even (B.29) is equivalent to a matrix factor model with no changes in the row column loadings space, and heteroskedastic factors. Hence, in both cases, it can be expected that our procedure should not detect any changes, since the space spanned by $\mathbf{R}_t$ does not change over time, and the presence of changes is absorbed by the common factors $\mathbf{F}_t$, implying that the form of time variation in (B.29) can be "tolerated" (as Bates et al., 2013 put it). The same considerations would apply if one were to consider the case where the matrix $\mathbf{R} = \mathbf{R}_t$ fluctuates around its average, $\mathbf{R}_0$, viz.

$$\mathbf{X}_t = \begin{cases} \mathbf{R}_0 \mathbf{F}_t \mathbf{C}' + \mathbf{E}_t, & 1 \leq t \leq m \\ \mathbf{R}_t \mathbf{F}_t \mathbf{C}' + \mathbf{E}_t, & m+1 \leq t \leq T \end{cases},$$

with $\mathbf{R}_t = \mathbf{R}_0 + \mathbf{U}_t$, and $\mathbf{U}_t$ is a zero mean, *i.i.d.* shock independent of $\{\mathbf{F}_t\}$ and $\{\mathbf{E}_t\}$, following the same assumptions as $\mathbf{E}_t$.

Conversely, our procedure is designed to have power in the presence of a change, denoted as $H_{A,4}$, like

$$\mathbf{X}_t = \begin{cases} \mathbf{R}_0 \mathbf{F}_t \mathbf{C}' + \mathbf{E}_t, & 1 \leq t \leq k^* \\ \mathbf{R}_0 \mathbf{F}_t \mathbf{C}' + \min\left\{1, \frac{t-k^*}{m}\right\} \widetilde{\mathbf{R}} \widetilde{\mathbf{F}}_t \mathbf{C}' + \mathbf{E}_t, & k^*+1 \leq t \leq T \end{cases}. \tag{B.31}$$

In Table B.14, we report the results of a small Monte Carlo under (B.27) and (B.31), with $\mathbf{R}_0$ and $\widetilde{\mathbf{R}}_t$ having entries generated as $i.i.d.\mathcal{U}(-1,1)$, and $\theta = 0.1$.

B.3.7. *On the use of vectorisation-based methods.* Finally, we have considered the performance of the methodology developed in Barigozzi and Trapani (2020), applied to $\text{Vec}(\mathbf{X}_t)$. The results in Table B.15 confirm that, when ensuring size control, essentially no power is found, thus making the case for the use of methodologies that are specifically designed for matrix-valued time series.

The results in Table B.15 are not unexpected, and we now offer an explanation as to why this is the case. When vectorising the matrix-valued data, the cross-sectional dimension of $\text{Vec}(\mathbf{X}_t)$ is equal to $p_1 p_2$. In the case of vectorised data, the theory developed in Trapani (2018) and Barigozzi

Table B.14

Results with smooth breaks, over 1000 replications, with significance level $\alpha = 0.05$, $k1 = k_2 = 3$, $m = p_2$, $r = 8$, $\epsilon = 0.05$, and $T = 200$. "PS", partial-sum; "WC", worst case.

			Matrix-Projection				Matrix-Without-Projection			
			PS			WC	PS			WC
T	p_1	p_2	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Empirical power ($\times 100$) under H_{A3}										
200	20	80	5.2	2.2	3.3	3.7	35.0	21.5	9.9	6.4
200	30	80	4.7	1.2	4.0	3.8	42.1	23.5	8.4	8.3
200	40	80	4.2	1.5	2.6	3.7	44.7	26.6	7.9	9.9
200	80	20	3.7	2.7	3.3	3.3	3.8	1.7	2.7	3.8
200	80	30	4.2	2.4	3.6	4.9	4.9	2.3	3.1	4.7
200	80	40	4.7	1.5	3.3	3.1	2.8	2.0	3.4	6.2
200	20	20	10.6	5.8	4.3	5.9	65.3	52.2	29.2	18.0
200	30	30	4.2	1.8	3.5	4.2	65.3	48.8	24.8	14.1
200	40	40	4.9	1.8	2.9	4.2	59.3	43.1	18.4	11.5
Empirical power ($\times 100$) under H_{A4}										
200	20	80	93.9	91.7	88.6	93.9	95.1	91.4	81.6	89.8
200	30	80	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0
200	40	80	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	80	20	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	80	30	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	80	40	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	20	20	99.6	99.4	99.0	99.2	100.0	99.9	99.3	99.0
200	30	30	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	40	40	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Empirical delay (interquantiles) under H_{A3}										
200	20	80	72,78,84	74,79,85	76,82,88	70,76,81	59,75,83	68,78,86	78,84,92	72,79,86
200	30	80	62,66,70	63,67,71	66,70,74	60,64,68	57,65,71	61,68,73	68,73,78	63,67,72
200	40	80	56,59,63	57,60,63	59,62,66	54,57,60	51,59,63	56,61,65	61,65,69	57,60,64
200	80	20	21,23,25	21,23,25	23,25,27	19,21,23	24,27,30	25,28,31	27,30,34	22,24,27
200	80	30	26,27,29	26,28,29	27,29,31	24,26,27	29,31,33	29,31,33	31,33,35	27,29,31
200	80	40	30,32,34	31,32,34	32,34,36	28,30,32	34,36,38	34,36,38	36,38,40	32,34,36
200	20	20	20,22,26	20,23,27	22,25,29	18,20,23	-20,19,24	-30,20,25	0,23,29	18,21,25
200	30	30	25,27,29	25,27,29	27,29,31	23,25,27	-3,24,28	-7,26,29	25,29,31	23,26,28
200	40	40	29,31,34	30,31,34	31,33,35	27,29,31	17,30,33	12,31,34	31,34,37	29,31,33

Table B.15

Results for the vectorisation approach, over 1000 replications, with significance level $\alpha = 0.05$, $k_1 = k_2 = 3$, $m = p_2$, $r = 8$, $\epsilon = 0.05$.

			Empirical size under H_0		Empirical power under H_{A1}		Empirical power under H_{A2}	
T	p_1	p_2	$\eta = 0.45$	$\eta = 0.5$	$\eta = 0.45$	$\eta = 0.5$	$\eta = 0.45$	$\eta = 0.5$
200	20	80	4.2	4.4	4.4	4.6	4.6	5.0
200	30	80	3.6	3.8	3.8	4.1	4.2	4.7
200	40	80	3.6	4.2	4.8	5.6	4.4	4.6
200	80	20	3.7	3.6	5.0	4.8	4.3	4.2
200	80	30	4.2	3.7	4.4	3.9	3.6	3.2
200	80	40	3.4	4.0	3.9	4.0	4.6	5.2
200	20	20	4.2	4.1	4.6	4.2	4.4	4.2
200	30	30	3.8	3.3	4.9	4.5	5.0	4.4
200	40	40	4.9	4.5	5.1	5.2	3.9	4.1
80	40	40	4.3	3.9	3.6	3.7	5.5	5.7
120	40	40	3.8	4.0	4.0	4.5	4.5	4.3
160	40	40	2.8	3.0	2.9	3.3	3.3	4.0
200	40	40	4.0	3.9	3.9	4.1	3.1	3.6
240	40	40	3.0	4.0	5.1	5.1	3.9	4.1
280	40	40	4.2	4.4	3.2	3.3	4.7	5.1
320	40	40	4.7	5.5	3.7	4.8	4.8	5.0
360	40	40	4.4	4.5	4.9	5.5	4.1	4.1
400	40	40	4.0	4.4	3.6	3.9	4.3	4.6

and Trapani (2020) can be applied directly, and it immediately yields that the spiked eigenvalues of the second moment matrix

$$\frac{1}{m}\sum_{t=1}^{m}\mathrm{Vec}\left(\mathbf{X}_t\right)\left(\mathrm{Vec}\left(\mathbf{X}_t\right)\right)',$$

diverge at a rate $\Omega_{a.s.}\left(p_1p_2\right)$ - this result is not surprising, since the leading eigenvalues of a second moment matrix, in the case of a (strong/pervasive) factor structure, diverge at a rate given by the cross-sectional dimension. Conversely, upon following verbatim the proof of Lemma 2 in Trapani (2018), it can be shown that the other eigenvalues are bounded by $o_{a.s.}\left(m^{-1/2}p_1p_2\right)$, where $\widetilde{l}_{p_1,p_2,m}$ is a slowly varying sequence. Intuitively, if p_1p_2 is very large, the discrepancy between spiked and non-spiked eigenvalues may not be large enough to be discernible. In order to implement our procedures, we need to rescale all eigenvalues by $\left(p_1p_2\right)^{-\delta^*}$, where δ^* is chosen (see e.g. Trapani, 2018) as

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

having defined $\beta^* = \ln\left(p_1p_2\right)/\ln m$. Hence, it is apparent that, when p_1p_2 is large relative to m, δ^* is also large (and larger than δ defined in (3.9)): upon comparing δ^* with δ, it emerges that p_2 "helps" when using our methodology, whereas it "hinders" when using a method for vectorised data. When using our methodology, under the alternative the building block of our monitoring scheme, i.e.

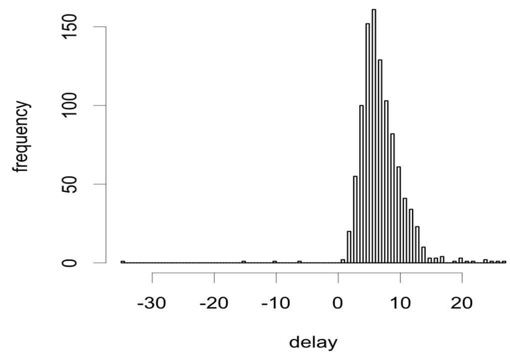
$$p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau},$$

diverges at a rate $\Omega_{a.s.}\left(p_1^{1-\delta}\right) = \Omega_{a.s.}\left(\left(mp_2\right)^{1/2-\varepsilon}\right)$. On the other hand, when using a vectorised approach, under the alternative $\widehat{\lambda}_{k_1+1,\tau}$ diverges at a rate $\Omega_{a.s.}\left(p_1p_2\right)$; hence, $\left(p_1p_2\right)^{-\delta^*}\widehat{\lambda}_{k_1+1,\tau}$ diverges at a rate $\Omega_{a.s.}\left(\left(p_1p_2\right)^{1-\delta^*}\right) = \Omega_{a.s.}\left(m^{1/2-\varepsilon}\right)$. This explains the lower (in fact, virtually non-existent) power of the vectorisation methodology.

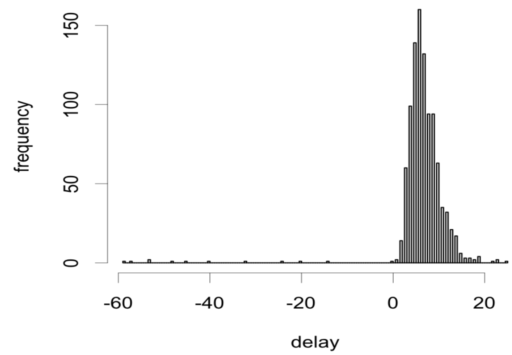
B.3.8. *Delay distribution and computational times.* In Figure B.1, we provide the histograms for the empirical delay of our statistics (using the projection-based estimation method) across simulations. As can be seen, there are few negative delays over the $1,000$ replications, indicating false detection. Interestingly, when using $\eta = 0.75$, false detections tend to concentrate at the beginning of the monitoring horizon; conversely, no obvious pattern is detected for other values of η.

Finally, in Table B.16, we report the average computational cost (measured in seconds, over $1,000$ replications) for our methodology based on using the projected method, the non-projected method, and the vectorisation method. In particular, we report only the computational times needed for the calculations of ψ_τ. As can be expected, in most cases, the projected method is more time-consuming, mainly because the projection matrix has to be re-estimated during the monitoring. The vectorisation approach is also computationally costly, mainly because the sample covariance matrix is "ultra" high-dimensional after vectorisation.

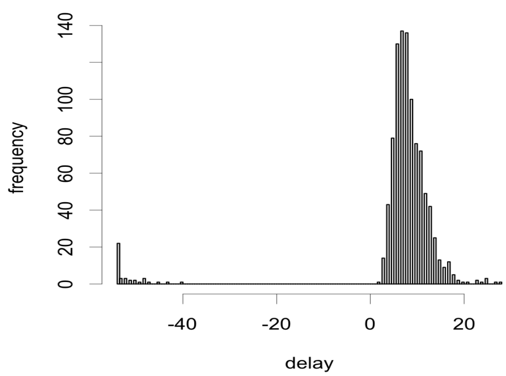
B.3.9. *Further empirical analysis.* We report further empirical evidence on Section 5 in the main paper, on monitoring the column space. We also report another application, to macro data.



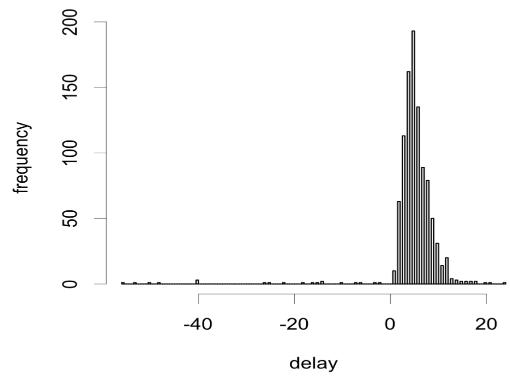
(a) "PS" ($\eta = 0.25$)



(b) "PS" ($\eta = 0.5$)



(c) "PS" ($\eta = 0.75$)



(a) "WC"

FIG B.1. *Histogram of the empirical delay under H_{A1} when $T = 200, p_1 = p_2 = 40$, over 1000 replications.*

Monitoring column factors for Fama and French 100 portfolios. We begin by investigating the effect of tuning parameters r, m, and ε on changepoint detection in the row factor structure of the data. Table B.17 reports the total number of change points (row factors), and the associated locations and types, detected by the projected worst-case statistic with different tuning parameters. According to Table B.17, the monitoring procedure is more robust to r and ε. Employing larger values of ε when checking if the factor space increases (and smaller values of ε when trying to detect if a factor has vanished) worsens the detection delay, similarly to increasing r; however, effects are minimal. Conversely, the choice of m is more sensitive, and a larger m will lead to longer detection delay and detection of fewer changepoints (only the "strong" ones). This is consistent with our remark in Section 4. In real examples, we will always suggest the users to try multiple values of m. The selection of m then depends on the users' tolerance of detection delay and their identification of influential breaks.

Turning to the column factor space, the rolling eigenvalue series of the row-row sample covariance matrices are plotted in Figure B.2. The trend is similar to that shown in Figure 1 of the main paper, but the variation of the second largest eigenvalue is less significant. Indeed, most of the test statistics fail to detect any change point for column factors if we still use the same tuning parameters as

Table B.16
Averaging computational cost (seconds) for the projected method (P), non-projected method (NP), and the vectorisation method (V), over 1000 replications.

		$T = 200$			$T = 300$			$T = 400$		
p_1	p_2	P	NP	V	P	NP	V	P	NP	V
20	80	2.31	0.12	2.95	4.03	0.22	5.35	5.63	0.30	7.53
30	80	3.38	0.17	4.28	6.08	0.30	7.96	8.42	0.41	11.28
40	80	5.39	0.24	5.85	9.41	0.42	10.74	13.00	0.58	15.29
80	20	4.72	0.40	0.55	7.13	0.60	0.86	9.31	0.80	1.14
80	30	5.89	0.43	1.66	8.93	0.65	2.56	12.14	0.87	3.54
80	40	7.55	0.46	3.41	11.66	0.71	5.44	15.69	0.94	7.48
20	20	0.68	0.13	0.18	1.01	0.19	0.26	1.36	0.26	0.35
30	30	1.37	0.15	0.69	2.13	0.24	1.06	2.95	0.33	1.47
40	40	2.69	0.20	1.74	4.12	0.31	2.75	5.65	0.42	3.82

Table B.17
Change points of ***row*** *factors for the Fama-French 100 portfolios with different tuning parameters, detected by the worst-case statistic with projection. (+) denotes the increase of factor number or change of loading space, while (-) denotes the decrease of factor number. We use ε_1 and ε_2 to denote the values of ε used in the construction of ψ_τ and $\widetilde{\psi}_\tau$ respectively.*

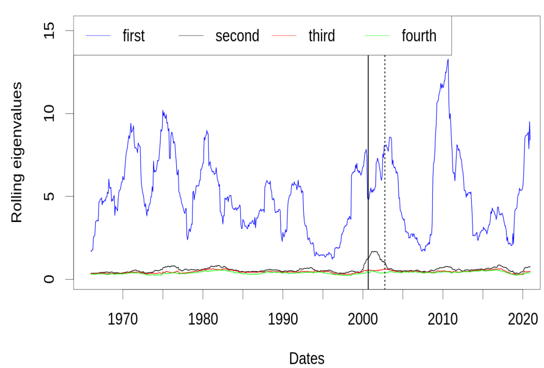
		$m = 24, \varepsilon_1 = 0.25, \varepsilon_2 = 0.05$				$r = 8, \varepsilon_1 = 0.25, \varepsilon_2 = 0.05$			
		$r = 4$	$r = 6$	$r = 10$	$r = 12$	$m = 18$	$m = 30$	$m = 36$	$m = 42$
Number of changes		4	4	4	4	4	3	2	0
Locations	First(+)	2000-08	2000-08	2000-07	2000-06	2000-02	2001-02	2001-08	NA
	Second(-)	2003-01	2002-10	2002-08	2002-07	2002-07	2003-09	2004-09	NA
	Third(+)	2014-06	2013-12	2013-10	2013-10	2013-04	2018-07	NA	NA
	Fourth(-)	2019-01	2018-12	2018-12	2018-12	2018-12	NA	NA	
		$r = 8, m = 24, \varepsilon_2 = 0.05$				$r = 8, m = 24, \varepsilon_1 = 0.25$			
		$\varepsilon_1 = 0.21$	$\varepsilon_1 = 0.23$	$\varepsilon_1 = 0.27$	$\varepsilon_1 = 0.29$	$\varepsilon_2 = 0.02$	$\varepsilon_2 = 0.04$	$\varepsilon_2 = 0.06$	$\varepsilon_2 = 0.08$
Number of changes		4	4	4	4	4	4	4	4
Locations	First(+)	2000-06	2000-06	2000-08	2000-08	2000-08	2000-07	2000-07	2000-07
	Second(-)	2002-07	2002-07	2002-10	2002-10	2002-12	2002-10	2002-08	2002-08
	Third(+)	2013-10	2013-10	2014-02	2014-08	2013-12	2013-11	2013-11	2013-11
	Fourth(-)	2018-12	2018-12	2018-12	2018-12	2019-01	2018-12	2018-12	2018-12

for the row factors. Therefore, there seems to be no sufficient evidence to assert that the column loadings change during the monitoring period. By way of robustness analysis, and to facilitate detection of "weaker" changes, we also reduce ε to 0.10 when constructing $\widetilde{\psi}_\tau$. Then, at most two change points are detected, as shown in Table B.18. The locations of change points by the projected worst-case statistic are plotted in Figure B.2, which also coincide with the variation of the second largest eigenvalue.

TABLE B.18
Location (year-month) of the change points for the Fama-French 100 portfolios. Parameter setting: $m = 24$, $r = 8$, $\varepsilon_2 = 0.05$, $\varepsilon_1 = 0.25$ for row factors and $\varepsilon_1 = 0.10$ for column factors. (+) denotes the increase of factor number or change of loading space, while (-) denotes the decrease of factor number.

		Matrix-Projection				Matrix-Without-Projection			
		PS			WC	PS			WC
	Change points	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Row	First(+)	2000-08	2000-09	2001-01	2000-07	2000-09	2000-12	NA	2000-08
	Second(-)	2002-11	2002-10	2003-07	2002-09	2003-06	2003-03	NA	2003-05
	Third(+)	2014-04	2014-06	2014-12	2013-11	2014-09	2015-05	NA	NA
	Fourth(-)	2019-01	2019-02	2019-02	2018-12	2019-05	2019-05	NA	NA
Column	First(+)	2000-09	2000-11	2001-07	2000-09	2001-02	2001-04	NA	2001-02
	Second(-)	2002-11	2002-12	2004-01	2002-10	2003-03	2003-05	NA	2003-03

Finally, we investigate the effects of tuning parameters using the worst-case statistic with projection. The results are reported in Table B.19, with similar conclusions are similar to those for the row factors.



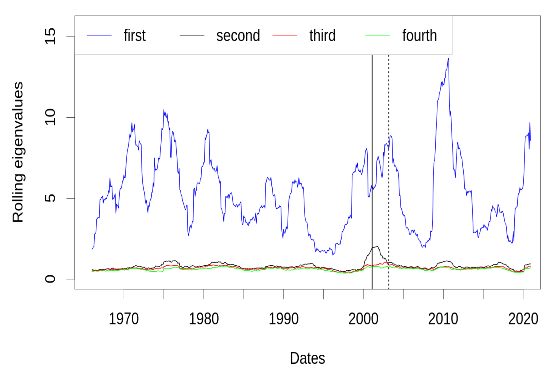


FIG B.2. *Eigenvalues of the rolling row-row sample covariance matrices with (left) and without (right) projection in the monitoring, for the Fama-French 100 portfolios. The real vertical lines are the locations when loading space changes or the number of factors increases, while the dashed vertical lines are the locations when the number of factors decreases, respectively.*

Multinational macroeconomic indices. In our second application, we study change point detection with macro data. We use the same data set as in Yu et al. (2022), which contains quarterly

TABLE B.19

*Change points of **column** factors for the Fama-French 100 portfolios with different tuning parameters, detected by the worst-case statistic with projection. (+) denotes the increase of factor number or change of loading space, while (-) denotes the decrease of factor number. We use ε_1 and ε_2 to denote the values of ε used in the construction of ψ_τ and $\widetilde{\psi}_\tau$ respectively.*

		$m = 24, \varepsilon_1 = 0.10, \varepsilon_2 = 0.05$				$r = 8, \varepsilon_1 = 0.10, \varepsilon_2 = 0.05$			
		$r = 4$	$r = 6$	$r = 10$	$r = 12$	$m = 18$	$m = 30$	$m = 36$	$m = 42$
Number of changes		2	2	2	2	4	2	0	0
Locations	First(+)	2000-02	2000-11	2000-08	2000-08	1993-01	2001-02	NA	NA
	Second(-)	2003-03	2002-11	2002-08	2002-08	1996-09	2003-09	NA	NA
	Third(+)	NA	NA	NA	NA	2000-07	NA	NA	NA
	Fourth(-)	NA	NA	NA	NA	2002-06	NA	NA	NA
		$r = 8, m = 24, \varepsilon_2 = 0.05$				$r = 8, m = 24, \varepsilon_1 = 0.10$			
		$\varepsilon_1 = 0.06$	$\varepsilon_1 = 0.08$	$\varepsilon_1 = 0.12$	$\varepsilon_1 = 0.14$	$\varepsilon_2 = 0.02$	$\varepsilon_2 = 0.04$	$\varepsilon_2 = 0.06$	$\varepsilon_2 = 0.08$
Number of changes		4	4	2	2	2	2	2	2
Locations	First(+)	1993-06	1994-07	2000-11	2001-02	2000-09	2000-09	2000-10	2000-09
	Second(-)	1996-10	1996-10	2002-12	2003-03	2002-10	2002-10	2002-11	2002-10
	Third(+)	2000-08	2000-08	NA	NA	NA	NA	NA	NA
	Fourth(-)	2002-09	2002-09	NA	NA	NA	NA	NA	NA

observations for 10 macroeconomic indices over 8 countries from 1988-Q1 to 2020-Q2.[7] Similar data sets have also been studied in Liu and Chen (2022) and Wang et al. (2019), which involves macroeconomic indices from more countries. We refer to Yu et al. (2022) for further details on the dataset, and the preprocessing steps; we removed outliers using the same approach as in the financial application.

The first step is to determine the numbers of row and column factors. In Figures B.3 and B.4 , we plot the leading five sample eigenvalues in the rolling monitoring process. Using the projected estimator in Yu et al. (2022) and the randomized testing procedures in He et al. (2023) suggests that $k_1 = 1$ or $k_1 = 2$ and $k_2 = 3$ or $k_2 = 4$, which essentially indicates that one row and one column factor may be weaker than the others. Combining this information with the eigenvalue gaps shown in Figures B.3 and B.4, we start with $k_1 = 2$ and $k_2 = 3$.

We use exactly the same methodology as in Section 5, although in this application we only consider the partial-sum statistics using $\eta = 0.25$ because the worst-case statistic reports no change points in most cases. After removing outliers, the test in Trapani (2016) suggests that at least 12 moments exist; hence, we use $r = 3$. We set $m = 20$ (5 years), and use the same values of ε as in Section 5 (we note that p_1 and p_2 are of comparable size across the two applications). The detected change points are reported in Table B.20. The partial-sum statistic reports at most two change points with the above tuning parameters. Combined with the rolling eigenvalue series, we believe that during the monitoring period, a row factor disappears first around "1999-Q2" and then a new row factor occurs around "2015-Q1". The results match with the variation of eigenvalues in Figure B.3. The second largest eigenvalue gradually decreases before "1999-Q2", while the relative size of the second largest eigenvalue over the first one gradually increases before "2015-Q1". For the column factors, a similar trend is found for the third largest eigenvalue around the two change

[7]The countries are the United States, the United Kingdom, Canada, France,Germany, Norway, Australia and New Zealand. The macroeconomic indices are from 4 groups, namely Consumer Price Index (CPI), Interest Rate (IR), Production (PRO) and International Trade (IT). The data can be freely downloaded from OECD data library.

points “2006-Q3” and “2015-Q1”. An interesting finding is that the change point locations of the macro data are not too far from those of the financial example.

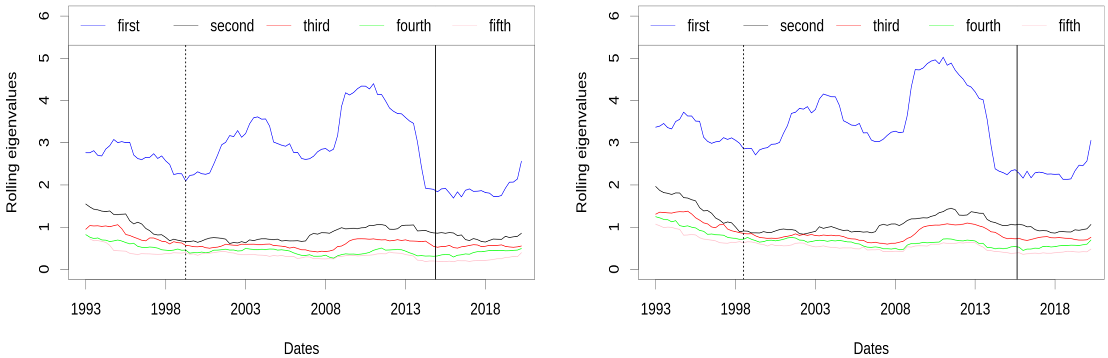


Fig B.3. *Eigenvalues of the rolling row-row sample covariance matrices with (left) and without (right) projection in the monitoring, for the macroeconomic indices. The real vertical lines are the locations when loading space changes or the number of factors increases, while the dashed vertical lines are the locations when the number of factors decreases, respectively.*

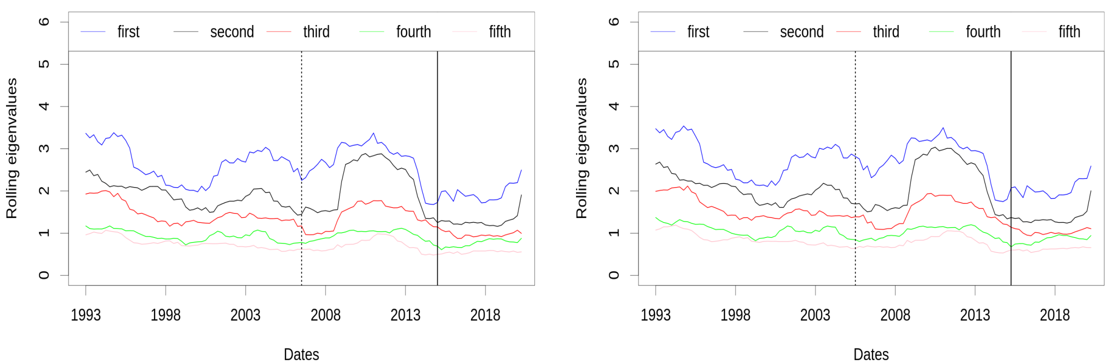


Fig B.4. *Eigenvalues of the rolling row-row sample covariance matrices with (left) and without (right) projection in the monitoring, for the macroeconomic indices. The real vertical lines are the locations when loading space changes or the number of factors increases, while the dashed vertical lines are the locations when the number of factors decreases, respectively.*

We also investigate the effects of tuning parameters to this data set, both for the row and column change points. The results are reported in Tables B.21 and B.22. The detected change points of the macro data are more sensitive to r and m, probably because the size and dimensions are smaller and the data are more heavy-tailed. Moreover, the partial-sum statistic is essentially less stable in small samples because it is more likely to lose control of size.

Table B.20

Location (year-quarter) of the change points for the macroeconomic indices. Parameter setting: $m = 20$, $r = 3$, $\epsilon_1 = 0.25$ and $\epsilon_2 = 0.05$. (+) denotes the increase of factor number or change of loading space, while (-) denotes the decrease of factor number.

		Matrix-Projection				Matrix-Without-Projection			
		PS			WC	PS			WC
	Change points	$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$		$\eta = 0.25$	$\eta = 0.5$	$\eta = 0.75$	
Row	First(-)	1999-Q2	1999-Q2	2000-Q2	NA	1998-Q3	1998-Q2	1999-Q1	2002-Q1
	Second(+)	2015-Q1	2015-Q4	NA	NA	2015-Q4	NA	NA	NA
Column	First(-)	2006-Q3	2007-Q3	NA	NA	2005-Q3	2006-Q4	NA	NA
	Second(+)	2015-Q1	2015-Q2	NA	NA	2015-Q2	2015-Q3	NA	NA

Table B.21

*Change points of **row** factors for the macroeconomic indices with different tuning parameters, detected by the partial-sum statistic with projection and $\eta = 0.25$. (+) denotes the increase of factor number or change of loading space, while (-) denotes the decrease of factor number.*

		$m = 20, \epsilon_1 = 0.25, \epsilon_2 = 0.05$				$r = 3, \epsilon_1 = 0.25, \epsilon_2 = 0.05$			
		$r = 2$	$r = 4$	$r = 6$	$r = 8$	$m = 10$	$m = 15$	$m = 25$	$m = 30$
Number of changes		2	2	2	2	2	2	2	2
Locations	First(-)	1999-Q1	2000-Q1	2000-Q4	2001-Q2	1998-Q2	1998-Q4	2000-Q2	2001-Q1
	Second(+)	2012-Q2	2015-Q4	NA	NA	2012-Q3	2013-Q4	2016-Q4	2017-Q2
	Third(-)	NA	NA	NA	NA	2017-03	NA	NA	NA
		$r = 3, m = 20, \epsilon_2 = 0.05$				$r = 3, m = 20, \epsilon_1 = 0.25$			
		$\epsilon_1 = 0.21$	$\epsilon_1 = 0.23$	$\epsilon_1 = 0.27$	$\epsilon_1 = 0.29$	$\epsilon_2 = 0.02$	$\epsilon_2 = 0.04$	$\epsilon_2 = 0.06$	$\epsilon_2 = 0.08$
Number of changes		2	2	2	2	2	2	2	2
Locations	First(-)	1999-Q3	1999-Q3	1999-Q2	1999-Q4	2000-Q2	1999-Q3	1999-Q2	1999-Q1
	Second(+)	2014-Q1	2014-Q2	2015-Q3	2016-Q2	2014-Q4	2014-Q4	2015-Q1	2015-Q1

Table B.22

*Change points of **column** factors for the macroeconomic indices with different tuning parameters, detected by the partial-sum statistic with projection and $\eta = 0.25$. (+) denotes the increase of factor number or change of loading space, while (-) denotes the decrease of factor number.*

		$m = 20, \epsilon_1 = 0.25, \epsilon_2 = 0.05$				$r = 3, \epsilon_1 = 0.25, \epsilon_2 = 0.05$			
		$r = 2$	$r = 4$	$r = 6$	$r = 8$	$m = 10$	$m = 15$	$m = 25$	$m = 30$
Number of changes		2	2	2	2	2	2	2	2
Locations	First(-)	2002-Q1	2008-Q4	2000-Q4	NA	1997-Q2	2005-Q1	2007-Q4	2008-Q4
	Second(+)	2012-Q1	NA	NA	NA	2010-Q2	2013-Q2	2017-Q3	NA
	Third(-)	NA	NA	NA	NA	2016-Q4	NA	NA	NA
		$r = 3, m = 20, \epsilon_2 = 0.05$				$r = 3, m = 20, \epsilon_1 = 0.25$			
		$\epsilon_1 = 0.21$	$\epsilon_1 = 0.23$	$\epsilon_1 = 0.27$	$\epsilon_1 = 0.29$	$\epsilon_2 = 0.02$	$\epsilon_2 = 0.04$	$\epsilon_2 = 0.06$	$\epsilon_2 = 0.08$
Number of changes		2	2	2	2	2	2	2	2
Locations	First(-)	2006-Q4	2006-Q1	2005-Q4	2006-Q3	2008-Q1	2006-Q1	2005-Q1	2002-Q2
	Second(+)	2014-Q3	2014-Q3	2015-Q3	2016-Q1	2015-Q4	2015-Q1	2014-Q4	2013-Q4

C. Technical lemmas. Recall that, given an $n_1 \times n_2$ matrix $\mathbf{A}$: the i-th row and the j-th column of $\mathbf{A}$ are defined as $\mathbf{A}_{i\cdot}$ and $\mathbf{A}_{\cdot j}$ respectively; $\|\mathbf{A}\|_{\max} = \max_{1\le i\le n_1}\max_{1\le j\le n_2}|A_{ij}|$; its Frobenius norm is denoted as $\|\mathbf{A}\|_F$; its Euclidean norm is denoted as $\|\mathbf{A}\|$. Moreover, we denote the (induced) L_q-norm, $q \ge 1$, as $\|\mathbf{A}\|_q$.

Our first lemma is used throughout the paper to derive almost sure bounds, and it is taken from He et al. (2023).

Lemma C.1. *Consider a multi-index partial sum process* $U_{S_1,..,S_h} = \sum_{i_2=1}^{S_2}\cdots\sum_{i_h=1}^{S_h}\xi_{i_1,...,i_h}$, *and assume that, for some* $q \ge 1$

$$E\sum_{i_1=1}^{S_1}|U_{S_1,..,S_h}|^q \le c_0 S_1 \prod_{j=2}^{h} S_j^{d_j},$$

where $d_j \ge 1$ *for all* $1 \le j \le h$. *Then it holds that*

$$\lim\sup_{\min\{S_1,...,S_h\}\to\infty} \frac{|U_{S_1,..,S_h}|}{S_1(\ln S_1)^{1+\epsilon}\prod_{j=2}^{h} S_j^{d_j/q}(\ln S_j)^{1+\frac{1}{q}+\epsilon}} = 0 \ a.s.,$$

for all $\epsilon > 0$.

Proof. The lemma follows from (minor modifications of the proof of) Lemma C.1 in He et al. (2023) □

The next three lemmas are taken from (with minor modifications) He et al. (2023), Barigozzi and Trapani (2020) and Yu et al. (2022) respectively. For all of them, it suffices to have $r = 2$ in Assumptions A.1-A.4.

Lemma C.2. *We assume that Assumptions A.1-A.4 are satisfied. Then, for all* $1 \le \tau \le T_m$, *it holds that there exist two constants* $0 < c_0 \le c_1 < \infty$ *such that* $p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau} = \Omega_{a.s.}(c_0)$ *and* $p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau} = O_{a.s.}(c_1)$.

Proof. The lemma is shown in He et al. (2023) - see their Lemma C.3 and its proof. □

Lemma C.3. *We assume that Assumptions A.1-A.5 are satisfied. Then it holds that there exist a triplet of random variables* $(p_{1,0}, p_{2,0}, m_0)$ *such that*

(i) *under (3.2)*

$$\widehat{\lambda}_{k_1+1,\tau}\begin{cases} \le c_0 & \text{for } \tau \le t^* \\ \ge c_1 \min\left\{\frac{\tau-t^*}{m}, \frac{m+t^*-\tau}{m}\right\} p_1 & \text{for } t^* < \tau < m+t^* \\ \le c_0 & \text{for } \tau \ge m+t^* \end{cases};$$

(ii) *under (3.4)*

$$\widehat{\lambda}_{k_1+1,\tau}\begin{cases} \le c_0 & \text{for } \tau \le t^* \\ \ge c_1 \frac{\tau-t^*}{m} p_1 & \text{for } t^* < \tau < m+t^* \\ \ge c_0 p_1 & \text{for } \tau \ge m+t^* \end{cases}.$$

PROOF. The proof of the lemma is based on very similar passages as the proof of Lemma 1 in Barigozzi and Trapani (2020), and thus we report only the main passages. We begin with part *(ii)* of the lemma, focusing on the case $t^* < \tau < m + t^*$. Note that

$$\begin{aligned}
\widehat{\lambda}_{k_1+1,\tau} &= \widehat{\lambda}_{k_1+1}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1}^{m+t^*}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t' + \frac{1}{p_2^2 m}\sum_{t=m+t^*}^{m+\tau}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t'\right) \\
&\geq \widehat{\lambda}_{\min}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1}^{m+t^*}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t'\right) + \widehat{\lambda}_{k_1+1}\left(\frac{1}{p_2^2 m}\sum_{t=m+t^*}^{m+\tau}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t'\right) \\
&\geq \widehat{\lambda}_{k_1+1}\left(\frac{1}{p_2^2 m}\sum_{t=m+t^*}^{m+\tau}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t'\right),
\end{aligned}$$

by Weyl's inequality. Also

$$\widehat{\lambda}_{k_1+1}\left(\frac{1}{p_2^2 m}\sum_{t=m+t^*}^{m+\tau}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t'\right) = \frac{\tau - t^*}{m}\widehat{\lambda}_{k_1+1}\left(\frac{1}{p_2^2\left(m+\tau-t^*\right)}\sum_{t=m+t^*}^{m+\tau}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t'\right);$$

the fact that

$$\widehat{\lambda}_{k_1+1}\left(\frac{1}{p_2^2\left(\tau-t^*\right)}\sum_{t=m+t^*}^{m+\tau}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t'\right) \geq c_0 p_1,$$

is an immediate consequence of Lemma C.1. The rest of the proof - for the cases $\tau \leq t^*$ and $\tau \geq m + t^*$ - follows immediately. We now turn to part (i) of the Lemma, and consider the case where all loadings change for simplicity - i.e. $\mathbf{R}_0$ is empty. In the interval $t^* < \tau < m + t^*$ we can write

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \end{bmatrix}\begin{bmatrix} \mathbf{F}_t I\left(t < t^*\right) & 0 \\ 0 & \mathbf{F}_t I\left(t \geq t^*\right) \end{bmatrix}\begin{bmatrix} \mathbf{C}' \\ \mathbf{C}' \end{bmatrix} + \mathbf{E}_t.$$

The second moment matrix has a component given by

$$\begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \end{bmatrix}\begin{bmatrix} \frac{1}{p_2^2 m}\sum_{t=\tau+1}^{m+t^*}\mathbf{F}_t I\left(t < t^*\right)\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t' & 0 \\ 0 & \frac{1}{p_2^2 m}\sum_{t=m+t^*}^{m+\tau}\mathbf{F}_t I\left(t \geq t^*\right)\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t' \end{bmatrix}\begin{bmatrix} \mathbf{R}_1' \\ \mathbf{R}_2' \end{bmatrix}$$

Using Theorem 7 in Merikoski and Kumar (2004)

$$\begin{aligned}
&\widehat{\lambda}_{k_1+1}\left(\begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \end{bmatrix}\begin{bmatrix} \frac{1}{p_2^2 m}\sum_{t=\tau+1}^{m+t^*}\mathbf{F}_t I\left(t < t^*\right)\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t' & 0 \\ 0 & \frac{1}{p_2^2 m}\sum_{t=m+t^*}^{m+\tau}\mathbf{F}_t I\left(t \geq t^*\right)\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t' \end{bmatrix}\begin{bmatrix} \mathbf{R}_1' \\ \mathbf{R}_2' \end{bmatrix}\right) \\
\geq\ &\widehat{\lambda}_{\min}\left(\begin{bmatrix} \frac{1}{p_2^2 m}\sum_{t=\tau+1}^{m+t^*}\mathbf{F}_t I\left(t < t^*\right)\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t' & 0 \\ 0 & \frac{1}{p_2^2 m}\sum_{t=m+t^*}^{m+\tau}\mathbf{F}_t I\left(t \geq t^*\right)\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t' \end{bmatrix}\right)\widehat{\lambda}_{k_1+1}\left(\begin{bmatrix} \mathbf{R}_1'\mathbf{R}_1 & \mathbf{R} \\ \mathbf{R}_2'\mathbf{R}_1 & \mathbf{R} \end{bmatrix}\right)
\end{aligned}$$

Using Lemma C.4 and standard passages, it can be shown that, as $\min\{p_1, p_2, m\} \to \infty$

$$\begin{aligned}
&\widehat{\lambda}_{\min}\left(\begin{bmatrix} \frac{1}{p_2^2 m}\sum_{t=\tau+1}^{m+t^*}\mathbf{F}_t I\left(t < t^*\right)\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t' & 0 \\ 0 & \frac{1}{p_2^2 m}\sum_{t=m+t^*}^{m+\tau}\mathbf{F}_t I\left(t \geq t^*\right)\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t' \end{bmatrix}\right) \\
=\ &\widehat{\lambda}_{\min}\left(\begin{bmatrix} \frac{m+t^*-\tau}{m}\mathbf{\Sigma}_1 & 0 \\ 0 & \frac{\tau-t^*}{m}\mathbf{\Sigma}_1 \end{bmatrix}\right) + o_{a.s.}(1),
\end{aligned}$$

where $\mathbf{\Sigma}_1$ is defined in Assumption A.1, with

$$\widehat{\lambda}_{\min}\left(\left[\begin{array}{cc} \frac{m+t^*-\tau}{m}\mathbf{\Sigma}_1 & 0 \\ 0 & \frac{\tau-t^*}{m}\mathbf{\Sigma}_1 \end{array}\right]\right) \geq c_0 \min\left\{\frac{m+t^*-\tau}{m}, \frac{\tau-t^*}{m}\right\}.$$

Also, Assumption A.5 readily entails that

$$\widehat{\lambda}_{k_1+1}\left(\left[\begin{array}{cc} \mathbf{R}_1'\mathbf{R}_1 & \mathbf{R}_1'\mathbf{R}_2 \\ \mathbf{R}_2'\mathbf{R}_1 & \mathbf{R}_2'\mathbf{R}_2 \end{array}\right]\right) \geq c_0 p_1.$$

Putting all together, it now follows that

$$\widehat{\lambda}_{k_1+1}\left(\frac{1}{m}\sum_{t=m+t^*}^{m+\tau} \widetilde{\mathbf{Y}}_t \widetilde{\mathbf{Y}}_t'\right) \geq c_0 p_1 \min\left\{\frac{m+t^*-\tau}{m}, \frac{\tau-t^*}{m}\right\},$$

for sufficiently large p_1, p_2, and m. □

LEMMA C.4. *We assume that Assumptions A.1-A.4 are satisfied. Then it holds that*

$$\left\|\widetilde{\mathbf{C}} - \mathbf{C}\mathbf{H}\right\|_F = o_{a.s.}\left(p_2^{1/2}\left(\frac{1}{p_2} + \frac{1}{\sqrt{Tp_1}}\right)\left(\ln m \ln^2 p_1 \ln^2 p_2\right)^{1/2+\epsilon}\right),$$

for all $\epsilon > 0$*, where*

$$\mathbf{H} = \left(\frac{1}{m}\sum_{t=1}^{m} \mathbf{F}_t'\mathbf{F}_t\right)\frac{\mathbf{C}'\widetilde{\mathbf{C}}}{p_2}\widehat{\mathbf{\Lambda}}_c^{-1}.$$

PROOF. The lemma is shown in He et al. (2023). □

We now present two novel lemmas, which are required to study the asymptotics under the null and under alternatives, respectively.

LEMMA C.5. *We assume that Assumptions A.1-A.4 are satisfied. Then, under the null, it holds that*

$$\sum_{\tau=1}^{T_m} \psi_\tau = o_{a.s.}(1).$$

PROOF. We show the theorem for the case $k_1 = k_2 = 1$, for simplicity and without loss of generality; also, in order to make the notation less burdensome, when possible we omit the matrix $\mathbf{H}$ in $\widetilde{\mathbf{C}} - \mathbf{C}\mathbf{H}$. Finally, we note that some passages in the proof of this lemma are the same as in the proof of Theorem 2 in He et al. (2023), which therefore are reported only briefly to save space.

Recall that

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

Lemma C.2 states that, for sufficiently large (p_1, p_2, m), there exists a $0 < c_0 < \infty$ such that

$$\left|\frac{p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}}{p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau}}\right|^r \leq c_0 p_1^{-r\delta}\left|\widehat{\lambda}_{k_1+1,\tau}\right|^r,$$

and therefore the proof of the lemma requires finding a bound for

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

We begin by noting that, by Weyl's inequality

$$\begin{aligned}
&\widehat{\lambda}_{k_1+1,\tau} \\
\le\ & \lambda_{k_1+1,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right) \\
&+\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'+\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'+\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right) \\
\le\ & \lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right)+\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right) \\
&+\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right) \\
=\ & \lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right)+2\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right),
\end{aligned}$$

having used the fact that, under the null, it holds that

$$\widehat{\lambda}_{k_1+1,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)=0,$$

and symmetry in the last passage. Using Minkowski's inequality, this means that we need to bound

$$p_1^{-r\delta}\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right)\right|^r, \tag{C.2}$$

$$p_1^{-r\delta}\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right)\right|^r. \tag{C.3}$$

We begin by studying (C.2); using again Minkowski's inequality

$$
\begin{aligned}
\text{(C.4)} \qquad & \sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right)\right|^r \\
\leq \;& c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\mathbf{C}\mathbf{C}'\mathbf{E}_t'\right)\right|^r \\
& +c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right)\right|^r \\
& +c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\widetilde{\mathbf{C}}'\mathbf{E}_t'\right)\right|^r \\
& +c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right)\right|^r \\
= \;& I+II+III+IV.
\end{aligned}
$$

Consider I, and note that

$$
\begin{aligned}
& \sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\mathbf{C}\mathbf{C}'\mathbf{E}_t'\right)\right|^r \\
\leq \;& c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}E\left(\mathbf{E}_t\mathbf{C}\mathbf{C}'\mathbf{E}_t'\right)\right)\right|^r \\
& +c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\mathbf{C}\mathbf{C}'\mathbf{E}_t'-E\left(\mathbf{E}_t\mathbf{C}\mathbf{C}'\mathbf{E}_t'\right)\right)\right|^r \\
= \;& I_a+I_b.
\end{aligned}
$$

He et al. (2023) show that (see the proof of their Theorem 2)

$$
\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\mathbf{C}\mathbf{C}'\mathbf{E}_t'\right)\leq c_0p_2^{-1},
$$

whence $I_a=O\left(T_mp_2^{-r}\right)$. Considering now I_b, note that

$$
\begin{aligned}
& \sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\mathbf{C}\mathbf{C}'\mathbf{E}_t'-E\left(\mathbf{E}_t\mathbf{C}\mathbf{C}'\mathbf{E}_t'\right)\right)\right|^r \\
\leq \;& \sum_{\tau=1}^{T_m}\left|\sum_{i,j=1}^{p_1}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\sum_{h,k=1}^{p_2}C_hC_k\left(e_{ih,t}e_{jk,t}-E\left(e_{ih,t}e_{jk,t}\right)\right)\right)^2\right|^{r/2},
\end{aligned}
$$

where C_h is the h-th element of the $p_2 \times 1$ vector $\mathbf{C}$, and

$$
\begin{aligned}
& E\left|\sum_{i,j=1}^{p_1}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\sum_{h,k=1}^{p_2} C_hC_k\left(e_{ih,t}e_{jk,t}-E\left(e_{ih,t}e_{jk,t}\right)\right)\right)^2\right|^{r/2} \\
\leq & \left(\max_{1\leq h,k\leq p_2}\left|C_h\right|^r\left|C_k\right|^r\right)p_1^{r-2}\sum_{i,j=1}^{p_1}p_2^{r-1}\sum_{h=1}^{p_2}E\left|\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\sum_{k=1}^{p_2}\left(e_{ih,t}e_{jk,t}-E\left(e_{ih,t}e_{jk,t}\right)\right)\right|^r \\
\leq & c_0\left(\frac{1}{mp_2^2}\right)^r p_1^{r-2}\sum_{i,j=1}^{p_1}p_2^{r-1}\sum_{h=1}^{p_2}E\left|\sum_{t=\tau+1-m}^{\tau}\sum_{k=1}^{p_2}\left(e_{ih,t}e_{jk,t}-E\left(e_{ih,t}e_{jk,t}\right)\right)\right|^r \\
\leq & c_0\left(\frac{1}{mp_2^2}\right)^r p_1^rp_2^r\left(Tp_2\right)^{r/2}\leq c_1\left(\frac{p_1}{m^{1/2}p_2^{1/2}}\right)^r
\end{aligned}
$$

having used Assumption A.2*(i)* on the third line, and Assumption A.3*(iii)(b)* on the last line. Using Lemma C.1 it follows that

$$
I_b = o_{a.s.}\left(T_m\left(\frac{p_1}{m^{1/2}p_2^{1/2}}\right)^r\right),
$$

for all $\epsilon > 0$, and therefore

$$
I = O\left(T_mp_2^{-r}\right)+o_{a.s.}\left(T_m\left(\frac{p_1}{m^{1/2}p_2^{1/2}}\right)^r\left(\ln T_m\right)^{1+\epsilon}\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{1+1/r+\epsilon}\right). \tag{C.5}
$$

We now turn to II. By the proof of Theorem 2 in He et al. (2023), we know that

$$
\begin{aligned}
& \sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right)\right|^r \\
\leq & \left\|\widetilde{\mathbf{C}}-\mathbf{C}\right\|_F^r\sum_{\tau=1}^{T_m}\left(\sum_{j=1}^{p_2}\left\|\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_{\cdot j,t}\mathbf{C}'\mathbf{E}_t'\right\|_F^2\right)^{r/2}.
\end{aligned}
$$

Hence, following the proof of Theorem 2 in He et al. (2023)

$$
\begin{aligned}
& \sum_{\tau=1}^{T_m}E\left(\sum_{j=1}^{p_2}\left\|\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_{\cdot j,t}\mathbf{C}'\mathbf{E}_t'\right\|_F^2\right)^{r/2} \\
\leq & \sum_{\tau=1}^{T_m}\left(\sum_{j=1}^{p_2}\left\|\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}E\left(\mathbf{E}_{\cdot j,t}\mathbf{C}'\mathbf{E}_t'\right)\right\|_F^2\right)^{r/2} \\
& +\sum_{\tau=1}^{T_m}E\left(\sum_{j=1}^{p_2}\left\|\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\left(\mathbf{E}_{\cdot j,t}\mathbf{C}'\mathbf{E}_t'-E\left(\mathbf{E}_{\cdot j,t}\mathbf{C}'\mathbf{E}_t'\right)\right)\right\|_F^2\right)^{r/2} \\
= & II_a+II_b.
\end{aligned}
$$

As far as II_a is concerned, it is easy to see that

$$
\begin{aligned}
& \sum_{\tau=1}^{T_m}\left(\sum_{j=1}^{p_2}\left\|\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}E\left(\mathbf{E}_{\cdot j,t}\mathbf{C}'\mathbf{E}_t'\right)\right\|_F^2\right)^{r/2} \\
\leq & \sum_{\tau=1}^{T_m}\left(\sum_{j=1}^{p_2}\left\|\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}E\left(\mathbf{E}_{\cdot j,t}\mathbf{C}'\mathbf{E}_t'\right)\right\|_1^2\right)^{r/2} \\
\leq & \sum_{\tau=1}^{T_m}p_2^{-3r/2-1}\sum_{j=1}^{p_2}\left(\max_{1\leq h\leq p_1}\left|\sum_{i=1}^{p_1}\sum_{k=1}^{p_2}C_kE\left(e_{ij,t}e_{hk,t}\right)\right|\right)^r \\
\leq & \max_{1\leq k\leq p_2}|C_k|^r\sum_{\tau=1}^{T_m}p_2^{-3r/2}\left(\max_{1\leq h\leq p_1}\sum_{i=1}^{p_1}\sum_{k=1}^{p_2}\left|E\left(e_{ij,t}e_{hk,t}\right)\right|\right)^r,
\end{aligned}
$$

whence Assumptions A.2*(i)* and A.3*(ii)(a)* yield

$$
II_a=O\left(T_mp_2^{-3r/2}\right).
$$

Also

$$
\begin{aligned}
& \sum_{\tau=1}^{T_m}E\left(\sum_{j=1}^{p_2}\left\|\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\left(\mathbf{E}_{\cdot j,t}\mathbf{C}'\mathbf{E}_t'-E\left(\mathbf{E}_{\cdot j,t}\mathbf{C}'\mathbf{E}_t'\right)\right)\right\|_F^2\right)^{r/2} \\
\leq & \sum_{\tau=1}^{T_m}\left(\frac{1}{mp_2^2}\right)^r p_2^{r/2-1}\sum_{j=1}^{p_2}E\left(\sum_{h,k=1}^{p_1}\left|\sum_{i=1}^{p_2}\sum_{t=\tau+1-m}^{\tau}C_i\left(e_{hj,t}e_{ki,t}-E\left(e_{hj,t}e_{ki,t}\right)\right)\right|^2\right)^{r/2} \\
\leq & \sum_{\tau=1}^{T_m}\max_{1\leq i\leq p_2}|C_i|^r\left(\frac{1}{mp_2^2}\right)^r p_1^{2(r/2-1)}p_2^{r/2-1}\sum_{h,k=1}^{p_1}\sum_{j=1}^{p_2}E\left|\sum_{i=1}^{p_2}\sum_{t=\tau+1-m}^{\tau}\left(e_{hj,t}e_{ki,t}-E\left(e_{hj,t}e_{ki,t}\right)\right)\right|^r \\
\leq & c_0T_m\left(\frac{1}{mp_2^2}\right)^r p_1^rp_2^{r/2}m^{r/2}p_2^{r/2}\leq c_1T_m\left(\frac{p_1}{m^{1/2}p_2}\right)^r,
\end{aligned}
$$

having used Assumptions A.2*(i)* and A.3*(iii)(b)* in the last line. Using now Lemmas C.1 and C.4 we have

$$
\begin{aligned}
& II_b \\
= & o_{a.s.}\left(T_mp_2^{-r}\left(\frac{1}{p_2}+\frac{1}{\sqrt{Tp_1}}\right)^r\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{r/2+\epsilon}\right) \\
& +o_{a.s.}\left(T_m\left(\frac{p_1}{m^{1/2}p_2^{1/2}}\right)^r\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)^r\left(\ln T_m\right)^{1+\epsilon}\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{1+r/2+1/r+\epsilon}\right).
\end{aligned}
$$

Hence it follows that

$$
\begin{aligned}
\text{(C.6)}\qquad & II \\
= & O\left(T_mp_2^{-3r/2}\right)+o_{a.s.}\left(T_mp_2^{-r}\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)^r\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{r/2+\epsilon}\right) \\
& +o_{a.s.}\left(T_m\left(\frac{p_1}{m^{1/2}p_2^{1/2}}\right)^r\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)^r\left(\ln T_m\right)^{1+\epsilon}\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{1+r/2+1/r+\epsilon}\right).
\end{aligned}
$$

The same can be shown for III in (C.4) by symmetry. Finally, we consider IV, studying

$$\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right)\right|^r.$$

We note that the element of the matrix $\left(mp_2^2\right)^{-1}\sum_{\tau+1-m}^{\tau}\mathbf{E}_t\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'$ in position h,k is

$$\begin{aligned}
&\left\{\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right\}_{h,k}\\
=\;&\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\left(\sum_{j=1}^{p_2}e_{hj,t}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_j\right)\left(\sum_{j=1}^{p_2}e_{kj,t}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_j\right)\\
=\;&\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\sum_{n',n=1}^{p_2}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_n\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_{n'}e_{hn,t}e_{kn',t}\\
=\;&\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\sum_{n',n=1}^{p_2}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_n\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_{n'}E\left(e_{hn,t}e_{kn',t}\right)\\
&+\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\sum_{n',n=1}^{p_2}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_n\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_{n'}\left(e_{hn,t}e_{kn',t}-E\left(e_{hn,t}e_{kn',t}\right)\right).
\end{aligned}$$

Therefore we study the bound

$$\begin{aligned}
&\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right)\right|^r\\
\leq\;&\sum_{\tau=1}^{T_m}\left|\frac{1}{mp_2^2}\max_{1\leq k\leq p_1}\sum_{h=1}^{p_1}\sum_{t=\tau+1-m}^{\tau}\sum_{n',n=1}^{p_2}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_n\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_{n'}\left|E\left(e_{hn,t}e_{kn',t}\right)\right|\right|^r\\
&+\sum_{\tau=1}^{T_m}\left|\sum_{h,k=1}^{p_1}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\sum_{n',n=1}^{p_2}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_n\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)_{n'}\left(e_{hn,t}e_{kn',t}-E\left(e_{hn,t}e_{kn',t}\right)\right)\right)^2\right|^{r/2}\\
=\;&IV_a+IV_b.
\end{aligned}$$

It is easy to see that Assumption A.3*(ii)*(b) implies that $\max_{1\leq k\leq p_1}\sum_{h=1}^{p_1}\left|E\left(e_{hn,t}e_{kn',t}\right)\right|<\infty$;

hence

$$
\begin{aligned}
IV_a &\leq c_0 \sum_{\tau=1}^{T_m} \left| \frac{m}{mp_2^2} \sum_{n',n=1}^{p_2} \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_n \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_{n'} \right|^r \\
&\leq c_0 \sum_{\tau=1}^{T_m} \left| \frac{1}{p_2^2} \sum_{n',n=1}^{p_2} \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_n \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_{n'} \right|^r \\
&= c_0 \sum_{\tau=1}^{T_m} \left| \frac{1}{p_2^2} \left(\sum_{n=1}^{p_2} \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_n \right)^2 \right|^r = c_0 T_m \left| \frac{1}{p_2^2} p_2 \sum_{n=1}^{p_2} \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_n^2 \right|^r \\
&= c_0 T_m \left(\frac{1}{p_2} \left\| \widetilde{\mathbf{C}} - \mathbf{C} \right\|_F^2 \right)^r = o_{a.s.} \left(T_m \left(\frac{1}{p_2} + \frac{1}{\sqrt{mp_1}} \right)^{2r} (\ln T_m)^{1+\epsilon} \left(\ln m \ln^2 p_1 \ln^2 p_2 \right)^{r/2+\epsilon} \right).
\end{aligned}
$$

Also, letting $\eta_{hknn',t} = e_{hn,t} e_{kn',t} - E\left(e_{hn,t} e_{kn',t} \right)$ for short, we study

$$
\begin{aligned}
& \left(\frac{1}{mp_2^2} \right)^r \left| \sum_{h,k=1}^{p_1} \left(\sum_{n',n=1}^{p_2} \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_n \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_{n'} \left(\sum_{t=\tau+1-m}^{\tau} \eta_{hknn',t} \right) \right)^2 \right|^{r/2} \\
\leq & \left(\frac{1}{mp_2^2} \right)^r \left| \sum_{h,k=1}^{p_1} \left\| \widetilde{\mathbf{C}} - \mathbf{C} \right\|_F^4 \sum_{n',n=1}^{p_2} \left(\sum_{t=\tau+1-m}^{\tau} \eta_{hknn',t} \right)^2 \right|^{r/2} \\
\leq & \left\| \widetilde{\mathbf{C}} - \mathbf{C} \right\|_F^{2r} \left(\frac{1}{mp_2^2} \right)^r (p_1 p_2)^{r-2} \sum_{h,k=1}^{p_1} \sum_{n',n=1}^{p_2} \left(\sum_{t=\tau+1-m}^{\tau} \eta_{hknn',t} \right)^r .
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& \sum_{\tau=1}^{T_m} \left| \sum_{h,k=1}^{p_1} \left(\frac{1}{mp_2^2} \sum_{t=\tau+1-m}^{\tau} \sum_{n',n=1}^{p_2} \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_n \left(\widetilde{\mathbf{C}} - \mathbf{C} \right)_{n'} \left(e_{hn,t} e_{kn',t} - E\left(e_{hn,t} e_{kn',t} \right) \right) \right)^2 \right|^{r/2} \\
\leq & \left\| \widetilde{\mathbf{C}} - \mathbf{C} \right\|_F^{2r} \sum_{\tau=1}^{T_m} \left(\frac{1}{mp_2^2} \right)^r (p_1 p_2)^{r-2} \sum_{h,k=1}^{p_1} \sum_{n',n=1}^{p_2} \left(\sum_{t=\tau+1-m}^{\tau} \eta_{hknn',t} \right)^r .
\end{aligned}
$$

Noting that, by Assumption A.3*(iii)(a)*

$$
\sum_{\tau=1}^{T_m} \left(\frac{1}{mp_2^2} \right)^r (p_1 p_2)^{r-2} \sum_{h,k=1}^{p_1} \sum_{n',n=1}^{p_2} E \left(\sum_{t=\tau+1-m}^{\tau} \eta_{hknn',t} \right)^r \leq c_0 T_m \left(\frac{p_1}{m^{1/2} p_2} \right)^r ,
$$

Lemmas C.1 and C.4 finally yield

$$
IV_b = o_{a.s.} \left(T_m \left(\frac{p_1}{m^{1/2}} \right)^r \left(\frac{1}{p_2} + \frac{1}{\sqrt{mp_1}} \right)^{2r} (\ln T_m)^{1+\epsilon} \left(\ln m \ln^2 p_1 \ln^2 p_2 \right)^{1+r+1/r+\epsilon} \right),
$$

so that

$$
\text{(C.7)} \qquad \begin{aligned} & IV \\ = \; & o_{a.s.}\left(T_m\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)^{2r}(\ln T_m)^{1+\epsilon}\left(\ln m \ln^2 p_1 \ln^2 p_2\right)^{r/2+\epsilon}\right) \\ & +o_{a.s.}\left(T_m\left(\frac{p_1}{m^{1/2}}\right)^{r}\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)^{2r}(\ln T_m)^{1+\epsilon}\left(\ln m \ln^2 p_1 \ln^2 p_2\right)^{1+r+1/r+\epsilon}\right). \end{aligned}
$$

Thus in (C.2) we have

$$
\text{(C.8)} \qquad \begin{aligned} & p_1^{-r\delta}\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right)\right|^r \\ = \; & O\left(T_m\left(\frac{1}{p_1^{\delta}p_2}\right)^{r}\right)+o_{a.s.}\left(T_m\left(\frac{p_1^{1-\delta}}{\sqrt{mp_2}}\right)^{r}(\ln T_m)^{1+\epsilon}\left(\ln m \ln^2 p_1 \ln^2 p_2\right)^{1+1/r+\epsilon}\right) \\ & +o_{a.s.}\left(p_1^{-r\delta}T_m\left(1+\frac{p_1}{m^{1/2}}\right)^{r}\left(\frac{1}{mp_1}\right)^{r}(\ln T_m)^{1+\epsilon}\left(\ln m \ln^2 p_1 \ln^2 p_2\right)^{r/2+\epsilon}\right). \end{aligned}
$$

We now turn to bounding (C.3), and study

$$
\text{(C.9)} \qquad \begin{aligned} & \sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right)\right|^r \\ \leq \; & c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{E}_t'\right)\right|^r \\ & +c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\mathbf{C}'\mathbf{E}_t'\right)\right|^r \\ & +c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right)\right|^r \\ & +c_0\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right)\right|^r \\ = \; & I+II+III+IV, \end{aligned}
$$

recalling that $\mathbf{C}'\mathbf{C}=p_2$. We begin by noting that, as far as I is concerned

$$
\begin{aligned} & \left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{E}_t'\right)\right|^r \\ \leq \; & \left(\frac{1}{mp_2}\right)^{r}\left|\sum_{i,j=1}^{p_1}\left(\sum_{t=\tau+1-m}^{\tau}\sum_{h=1}^{p_2}R_iC_h\mathbf{F}_te_{jh,t}\right)^2\right|^{r/2} \\ \leq \; & \max_{1\leq i\leq p_1}|R_i|^r\max_{1\leq h\leq p_2}|C_h|^r\left(\frac{1}{mp_2}\right)^{r}p_1^{r-2}\sum_{i,j=1}^{p_1}\left(\sum_{t=\tau+1-m}^{\tau}\sum_{h=1}^{p_2}\mathbf{F}_te_{jh,t}\right)^{r}. \end{aligned}
$$

Then, using Assumptions A.2*(i)* and A.4*(i)(b)*

$$E\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{E}_t'\right)\right|^r\leq c_0T_m\left(\frac{p_1}{m^{1/2}p_2^{1/2}}\right)^r,$$

which, by Lemma C.1, yields that in (C.9)

$$I=o_{a.s.}\left(T_m\left(\frac{p_1}{m^{1/2}p_2^{1/2}}\right)^r(\ln T_m)^{1+\epsilon}\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{1+1/r+\epsilon}\right). \tag{C.10}$$

Consider now II in (C.9). Following the passages in the proof of Theorem 2 in He et al. (2023), it can be shown that

$$\begin{aligned}
&\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\mathbf{C}'\mathbf{E}_t'\right)\right|^r\\
\leq\ &\left(\frac{1}{mp_2^2}\right)^r\left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\mathbf{C}'\mathbf{E}_t'\right\|_F^r\\
\leq\ &\left(\frac{1}{mp_2^2}\right)^r\left|\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\right|^r\left|\sum_{i,k=1}^{p_1}\left(\sum_{h=1}^{p_2}\sum_{t=\tau+1-m}^{\tau}R_iC_h\mathbf{F}_te_{kh,t}\right)^2\right|^{r/2}\\
\leq\ &\left(\max_{1\leq i\leq p_1}|R_i|^r\right)\left(\max_{1\leq h\leq p_2}|C_h|^r\right)\left(\frac{1}{mp_2^2}\right)^r\|\mathbf{C}\|^r\left\|\widetilde{\mathbf{C}}-\mathbf{C}\right\|_F^r p_1^{r-2}\sum_{i,k=1}^{p_1}\left|\sum_{h=1}^{p_2}\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_te_{kh,t}\right|^r,
\end{aligned}$$

so that, by Assumptions A.2*(i)*-*(ii)*, it follows that

$$II\leq c_0\left(\frac{1}{mp_2^2}\right)^r\left\|\widetilde{\mathbf{C}}-\mathbf{C}\right\|_F^r p_1^{r-2}\sum_{\tau=1}^{T_m}\sum_{i,k=1}^{p_1}\left|\sum_{h=1}^{p_2}\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_te_{kh,t}\right|^r.$$

Using Assumption A.4*(i)(b)*, and Lemmas C.1 and C.4, it finally follows that

$$II=o_{a.s.}\left(T_m\left(\frac{p_1}{\sqrt{mp_2}}\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)\right)^r(\ln T_m)^{1+\epsilon}\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{1+r/2+1/r+\epsilon}\right). \tag{C.11}$$

We now turn to studying III in (C.9); passages are similar to the proof of Lemma B.3 in Yu et al.

(2022), and therefore we omit them when possible. It holds that

$$\begin{aligned}
& \left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{RF}_t\mathbf{C}'\mathbf{C}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right)\right|^r \\
= & \left|\lambda_{\max,\tau}\left(\frac{1}{mp_2}\sum_{t=\tau+1-m}^{\tau}\mathbf{RF}_t\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right)\right|^r \\
\leq & \left|\widehat{\mathbf{\Lambda}}^{-1}\right|^r\|\mathbf{R}\|_F^r\left\|\frac{1}{mp_2}\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\left(\frac{1}{mp_1p_2}\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\mathbf{E}_t'\mathbf{RF}_t\right)\mathbf{C}'\mathbf{E}_t'\right\|_F^r \\
& +\left|\widehat{\mathbf{\Lambda}}^{-1}\right|^r\|\mathbf{R}\|_F^r\left\|\frac{1}{mp_2}\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\left(\frac{1}{mp_1p_2}\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\mathbf{E}_t\right)\mathbf{E}_t'\right\|_F^r \\
& +\left|\widehat{\mathbf{\Lambda}}^{-1}\right|^r\|\mathbf{R}\|_F^r\left\|\frac{1}{mp_2}\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\left(\frac{1}{mp_1p_2}\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\mathbf{E}_t'\mathbf{E}_t\right)\mathbf{E}_t'\right\|_F^r \\
= & III_a+III_b+III_c,
\end{aligned}$$

where $\widehat{\mathbf{\Lambda}}$ is a diagonal matrix containing the k_1 largest eigenvalues of $(Tp_2)^{-1}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t'$. Following the proof of Theorem 1 in He et al. (2023), it follows immediately that $\left|\widehat{\mathbf{\Lambda}}^{-1}\right|^r = O_{a.s.}(1)$. Assumption A.2*(ii)* also implies that $\|\mathbf{R}\|_F^r = c_0(p_1)^{r/2}$. Also, it holds that

$$\begin{aligned}
& \left\|\frac{1}{mp_2}\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\left(\frac{1}{mp_1p_2}\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\mathbf{E}_t'\mathbf{RF}_t\right)\mathbf{C}'\mathbf{E}_t'\right\|_F^r \\
\leq & \left(\frac{1}{m^2p_1p_2^2}\right)^r\left\|\sum_{t=1}^{m}\mathbf{C}'\mathbf{E}_t'\mathbf{RF}_t\right\|_F^r\left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{C}'\mathbf{E}_t'\right\|_F^r \\
& +\left(\frac{1}{m^2p_1p_2^2}\right)^r\left\|\sum_{t=1}^{m}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\mathbf{RF}_t\right\|_F^r\left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{C}'\mathbf{E}_t'\right\|_F^r \\
= & III_{a,1}+III_{a,2}.
\end{aligned}$$

It holds that

$$\begin{aligned}
& E\sum_{\tau=1}^{T_m}\left(\frac{1}{m^2p_1p_2^2}\right)^r\left\|\sum_{t=1}^{m}\mathbf{C}'\mathbf{E}_t'\mathbf{RF}_t\right\|_F^r\left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{C}'\mathbf{E}_t'\right\|_F^r \\
\leq & \left(\frac{1}{m^2p_1p_2^2}\right)^r\sum_{\tau=1}^{T_m}\left(E\left\|\sum_{t=1}^{m}\mathbf{C}'\mathbf{E}_t'\mathbf{RF}_t\right\|_F^{2r}\right)^{1/2}\left(E\left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{C}'\mathbf{E}_t'\right\|_F^{2r}\right)^{1/2} \\
\leq & c_0T_m\left(\frac{1}{m^2p_1p_2^2}\right)^r(mp_1p_2)^r = c_0T_m\left(\frac{1}{mp_2}\right)^r,
\end{aligned}$$

where the last passage can be shown by explicitly computing the norms and using Assumptions

A.2*(i)*, A.4*(i)(b)* and A.4*(i)(d)*. Also

$$\begin{aligned} & \left(\frac{1}{m^2p_1p_2^2}\right)^r \left\|\sum_{t=1}^{m}\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\mathbf{R}\mathbf{F}_t\right\|_F^r \left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{C}'\mathbf{E}_t'\right\|_F^r \\ \leq & \left(\frac{1}{m^2p_1p_2^2}\right)^r \left\|\widetilde{\mathbf{C}}-\mathbf{C}\right\|_F^r \left\|\sum_{t=1}^{m}\mathbf{E}_t'\mathbf{R}\mathbf{F}_t\right\|_F^r \left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{C}'\mathbf{E}_t'\right\|_F^r ; \end{aligned}$$

hence we estimate

$$\begin{aligned} & \left(\frac{1}{m^2p_1p_2^2}\right)^r \left\|\widetilde{\mathbf{C}}-\mathbf{C}\right\|_F^r E\sum_{\tau=1}^{T_m}\left\|\sum_{t=1}^{m}\mathbf{E}_t'\mathbf{R}\mathbf{F}_t\right\|_F^r \left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{C}'\mathbf{E}_t'\right\|_F^r \\ \leq & \left(\frac{1}{m^2p_1p_2^2}\right)^r \left\|\widetilde{\mathbf{C}}-\mathbf{C}\right\|_F^r \sum_{\tau=1}^{T_m}\left(E\left\|\sum_{t=1}^{m}\mathbf{E}_t'\mathbf{R}\mathbf{F}_t\right\|_F^{2r}\right)^{1/2} \left(E\left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{C}'\mathbf{E}_t'\right\|_F^{2r}\right)^{1/2} \\ \leq & c_0\left\|\widetilde{\mathbf{C}}-\mathbf{C}\right\|_F^p T_m\left(\frac{1}{mp_2}\right)^p, \end{aligned}$$

having used Assumptions A.2*(i)* and A.4*(i)(b)*-*(c)*. Using now Lemmas C.1 and C.4 it follows that

$$\begin{aligned} III_{a,1} &= o_{a.s.}\left(T_m\left(\frac{1}{mp_2}\right)^r(\ln T_m)^{1+\epsilon}\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{1+1/r+\epsilon}\right), \\ III_{a,2} &= o_{a.s.}\left(T_m\left(\frac{1}{mp_2}\right)^r\left(p_2^{1/2}\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)\right)^r(\ln T_m)^{1+\epsilon}\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{1+r/2+1/r+\epsilon}\right), \end{aligned}$$

whence finally

$$III_a = o_{a.s.}\left(T_m\left(\frac{p_1^{1/2}}{mp_2}\right)^r\left(1+\sqrt{\frac{p_2}{mp_1}}\right)^r(\ln T_m)^{1+\epsilon}\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{1+r/2+1/r+\epsilon}\right).$$

We now consider

$$\begin{aligned} & \left\|\frac{1}{m^2p_1p_2^2}\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\mathbf{E}_t\right)\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right)\right\|_F^r \\ \leq & \left(\frac{1}{m^2p_1p_2^2}\right)^r\left\|\widetilde{\mathbf{C}}'\mathbf{C}\right\|_F^r\left\|\left(\sum_{t=1}^{m}\mathbf{F}_t'\mathbf{R}'\mathbf{E}_t\right)\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right)\right\|_F^r \\ \leq & c_0\left(\frac{1}{m^2p_1p_2}\right)^r\left\|\sum_{t,s=1}^{m}\mathbf{F}_t'\mathbf{R}'\mathbf{E}_t\mathbf{F}_{s+\tau-m}\mathbf{E}_{s+\tau-m}'\right\|_F^r \\ = & c_0\left(\frac{1}{m^2p_1p_2}\right)^r\left|\sum_{j=1}^{p_1}\left(\sum_{i=1}^{p_1}\sum_{h=1}^{p_2}\sum_{t,s=1}^{m}R_i\mathbf{F}_t\mathbf{F}_{s+\tau-m}e_{ih,t}e_{jh,s+\tau-m}\right)^2\right|^{r/2} \\ \leq & c_0\left(\frac{1}{m^2p_1p_2}\right)^r p_1^{r/2-1}\sum_{j=1}^{p_1}\left|\sum_{i=1}^{p_1}\sum_{h=1}^{p_2}\sum_{t,s=1}^{m}R_i\mathbf{F}_t\mathbf{F}_{s+\tau-m}e_{ih,t}e_{jh,s+\tau-m}\right|^r. \end{aligned}$$

Using the notation (see Assumption A.4*(ii)*) $\overline{\zeta}_{ih} = m^{-1/2}\sum_{t=1}^{m}\mathbf{F}_t e_{ih,t}$ and $\overline{\zeta}_{jh} = m^{-1/2}\sum_{s=1}^{m}\mathbf{F}_{s+\tau-m} e_{jh,s+\tau-m}$, we now estimate

$$
\begin{aligned}
& \left(\frac{1}{mp_1p_2}\right)^r \sum_{\tau=1}^{T_m} p_1^{r/2-1} \sum_{j=1}^{p_1} E\left|\sum_{i=1}^{p_1}\sum_{h=1}^{p_2} R_i \overline{\zeta}_{ih}\overline{\zeta}_{jh}\right|^r \\
\leq\ & c_0 T_m \left(\frac{p_1^{1/2}}{mp_1p_2}\right)^r E\left|\sum_{i=1}^{p_1}\sum_{h=1}^{p_2} R_i \overline{\zeta}_{ih}\overline{\zeta}_{jh}\right|^r \\
\leq\ & c_0 T_m \left(\frac{p_1^{1/2}}{mp_1p_2}\right)^r \left|\sum_{i=1}^{p_1}\sum_{h=1}^{p_2} R_i E\left(\overline{\zeta}_{ih}\overline{\zeta}_{jh}\right)\right|^r \\
& + c_0 T_m \left(\frac{p_1^{1/2}}{mp_1p_2}\right)^r E\left|\sum_{i=1}^{p_1}\sum_{h=1}^{p_2} R_i \left(\overline{\zeta}_{ih}\overline{\zeta}_{jh} - E\left(\overline{\zeta}_{ih}\overline{\zeta}_{jh}\right)\right)\right|^r \\
=\ & III_{b,1} + III_{b,2}.
\end{aligned}
$$

Assumption A.4*(ii)(a)* immediately entails that $III_{b,1} = O\left(T_m\left(mp_1^{1/2}\right)^{-r}\right)$; using Assumption A.4*(ii)(c)*, it also follows that

$$
III_{b,2} = o_{a.s.}\left(T_m\left(\frac{1}{mp_2^{1/2}}\right)^r (\ln T_m)^{1+\epsilon}\left(\ln m \ln^2 p_1 \ln^2 p_2\right)^{1+r/2+1/r+\epsilon}\right),
$$

so that

$$
III_b = O\left(T_m m^{-r}\right) + o_{a.s.}\left(T_m\left(\frac{p_1^{1/2}}{mp_2^{1/2}}\right)^r (\ln T_m)^{1+\epsilon}\left(\ln m \ln^2 p_1 \ln^2 p_2\right)^{1+r/2+1/r+\epsilon}\right).
$$

Finally, considering III_c

$$
\begin{aligned}
& \lambda_{\max,\tau}\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\mathbf{E}_t'\mathbf{E}_t\right)\mathbf{E}_t'\right) \\
=\ & \lambda_{\max,\tau}\left(\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\mathbf{E}_t'\mathbf{E}_t\right)\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right)\right) \\
\leq\ & \lambda_{\max,\tau}\left(\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}' E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right)\right) \\
& + \lambda_{\max,\tau}\left(\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\left(\mathbf{E}_t'\mathbf{E}_t - E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\right)\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right)\right) \\
=\ & III_{c,1} + III_{c,2}.
\end{aligned}
$$

We have, by the multiplicative Weyl's inequality (see Theorem 7 in Merikoski and Kumar, 2004)

$$\begin{aligned}
& \lambda_{\max,\tau}\left(\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}'E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right)\right) \\
\leq\ & \lambda_{\max,\tau}\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}'E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\lambda_{\max,\tau}\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right) \\
\leq\ & \left\|\widetilde{\mathbf{C}}\right\|_{\mathbf{F}}\lambda_{\max,\tau}\left(\sum_{t=1}^{m}E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\lambda_{\max,\tau}\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right) \\
\leq\ & p_2^{1/2}\lambda_{\max,\tau}\left(\sum_{t=1}^{m}E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\lambda_{\max,\tau}\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right),
\end{aligned}$$

so that, by similar passages as above

$$\begin{aligned}
& \left(\frac{1}{m^2p_1p_2^2}\right)^r\sum_{\tau=1}^{T_m}E\left|\lambda_{\max,\tau}\left(\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}'E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right)\right)\right|^r \\
\leq\ & \left(\frac{1}{m^2p_1p_2^2}\right)^r\sum_{\tau=1}^{T_m}p_2^{r/2}\left(E\left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right\|_F^{2r}\right)^{1/2}\left(\left|\lambda_{\max,\tau}\left(\sum_{t=1}^{m}E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\right|^{2r}\right)^{1/2} \\
\leq\ & c_0T_m\left(\frac{p_1^{1/2}}{m^{1/2}p_2}\right)^r,
\end{aligned}$$

having used the fact that

$$\lambda_{\max,\tau}\left(\sum_{t=1}^{m}E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\leq c_0mp_1,$$

which is shown in the proof of Lemma C.1 in He et al. (2023). Hence, by Lemma C.1

$$III_{c,1}=o_{a.s.}\left(T_m\left(\frac{p_1^{1/2}}{m^{1/2}p_2}\right)^r\left(\ln T_m\right)^{1+\epsilon}\left(\ln^2 m\ln^3 p_1\ln^3 p_2\right)^{1+1/r+\epsilon}\right).$$

Similarly

$$\begin{aligned}
& \lambda_{\max,\tau}\left(\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\left(\mathbf{E}_t'\mathbf{E}_t-E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\right)\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right)\right) \\
\leq\ & p_2^{1/2}\left\|\sum_{t=1}^{m}\left(\mathbf{E}_t'\mathbf{E}_t-E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\right\|_F\left\|\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right\|_F,
\end{aligned}$$

so that

$$
\begin{aligned}
&\left(\frac{1}{m^2p_1p_2^2}\right)^r\sum_{\tau=1}^{T_m}E\left|\lambda_{\max,\tau}\left(\left(\sum_{t=1}^{m}\widetilde{\mathbf{C}}'\left(\mathbf{E}_t'\mathbf{E}_t-E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\right)\left(\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right)\right)\right|^r\\
\leq\;&\left(\frac{1}{m^2p_1p_2^{3/2}}\right)^r\sum_{\tau=1}^{T_m}\left(E\left\Vert\sum_{t=1}^{m}\left(\mathbf{E}_t'\mathbf{E}_t-E\left(\mathbf{E}_t'\mathbf{E}_t\right)\right)\right\Vert_F^{2r}\right)^{1/2}\left(E\left\Vert\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\mathbf{E}_t'\right\Vert_F^{2r}\right)^{1/2}\\
\leq\;&c_0T_m\left(\frac{p_1^{1/2}}{mp_2^{1/2}}\right)^r,
\end{aligned}
$$

having used again Assumptions A.3*(iii)(c)* and A.4*(i)(a)*. Putting all together

$$
\begin{aligned}
III_c\;=\;&o_{a.s.}\left(T_m\left(\frac{p_1^{1/2}}{m^{1/2}p_2}\right)^r(\ln T_m)^{1+\epsilon}\left(\ln^2 m\ln^3 p_1\ln^3 p_2\right)^{1+1/r+\epsilon}\right)\\
&+o_{a.s.}\left(T_m\left(\frac{p_1^{1/2}}{mp_2^{1/2}}\right)^r(\ln T_m)^{1+\epsilon}\left(\ln^2 m\ln^3 p_1\ln^3 p_2\right)^{1+1/r+\epsilon}\right),
\end{aligned}
$$

whence, finally

$$
\begin{aligned}
&III \qquad\qquad (C.12)\\
=\;&o_{a.s.}\left(T_m\left(\frac{p_1}{m^{1/2}p_2}\right)^r(\ln T_m)^{1+\epsilon}\left(\ln^2 m\ln^3 p_1\ln^3 p_2\right)^{1+1/r+\epsilon}\right)\\
&+o_{a.s.}\left(T_m\left(\frac{p_1}{mp_2^{1/2}}\right)^r(\ln T_m)^{1+\epsilon}\left(\ln^2 m\ln^3 p_1\ln^3 p_2\right)^{1+1/r+\epsilon}\right)+O\left(T_m\left(\frac{1}{m}\right)^r\right)\\
&+o_{a.s.}\left(T_m\left(\frac{p_1}{mp_2}\right)^r\left(1+\sqrt{\frac{p_2}{mp_1}}\right)(\ln T_m)^{1+\epsilon}\left(\ln^2 m\ln^3 p_1\ln^3 p_2\right)^{1+1/r+\epsilon}\right)
\end{aligned}
$$

We now consider term IV in (C.9), namely we study

$$
\begin{aligned}
&\sum_{\tau=1}^{T_m}\left\Vert\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right\Vert_F^r\\
\leq\;&\Vert\mathbf{R}\Vert_F^r\Vert\mathbf{C}\Vert_F^r\left\Vert\widetilde{\mathbf{C}}-\mathbf{C}\right\Vert_F^r\sum_{\tau=1}^{T_m}\left\Vert\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{E}_t'\right\Vert_F^r;
\end{aligned}
$$

hence it is easy to see, by the same passages as above, that this is dominated by III. Thus in (C.3) we have

$$
\begin{aligned}
&p_1^{-r\delta}\sum_{\tau=1}^{T_m}\left|\lambda_{\max,\tau}\left(\frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t'\right)\right|^r \qquad (C.13)\\
=\;&o_{a.s.}\left(T_m\left(\frac{p_1^{1-\delta}}{\sqrt{mp_2}}\right)^r(\ln T_m)^{1+\epsilon}\left(\ln m\ln^2 p_1\ln^2 p_2\right)^{1+1/r+\epsilon}\right)+O\left(T_m\left(\frac{1}{p_1^{\delta}m}\right)^r\right)
\end{aligned}
$$

The desired result now follows from putting (C.8) and (C.13) together, whence it ultimately follows that

$$
\begin{aligned}
& \sum_{\tau=1}^{T_m} \psi_\tau \\
= \; & O\left(T_m\left(\frac{1}{p_1^{\delta} p_2}\right)^r\right)+O\left(T_m\left(\frac{1}{p_1^{\delta} m}\right)^r\right) \\
& +o_{a.s.}\left(T_m\left(\frac{p_1^{1-\delta}}{\sqrt{m p_2}}\right)^r\left(\ln T_m\right)^{1+\epsilon}\left(\ln m \ln^2 p_1 \ln^2 p_2\right)^{1+1/r+\epsilon}\right),
\end{aligned}
$$

and using (3.12). □

Lemma C.6. *We assume that Assumptions A.1-A.5 are satisfied. Then, under (3.2), it holds that, for all $t^*<j$, there are two numbers $0<a<b<1$ such that*

$$
\sum_{\tau=t^*+am}^{t^*+m-bm}\left|\frac{p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}}{p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau}}\right|^r = c_0 m p_1^{r(1-\delta)}+o_{a.s.}\left(m p_1^{r(1-\delta)}\right). \tag{C.14}
$$

Under (3.4), it holds that for all $t^*<j$, there is a number $0<a$ such that

$$
\sum_{\tau=t^*+am}^{T_m}\left|\frac{p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}}{p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau}}\right|^r = c_0 T_m p_1^{r(1-\delta)}+o_{a.s.}\left(T_m p_1^{r(1-\delta)}\right). \tag{C.15}
$$

Proof. We focus only on (C.14), the proof of (C.15) being similar and, in fact, easier. We begin by noting that, by Lemma C.2, we have

$$
\left|\frac{p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}}{p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau}}\right|^r \geq c_0\left|p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}\right|^r,
$$

for some $c_0>0$, so that ultimately

$$
\sum_{\tau=t^*+1}^{j}\left|\frac{p_1^{-\delta}\widehat{\lambda}_{k_1+1,\tau}}{p_1^{-1}\sum_{j=1}^{p_1}\widehat{\lambda}_{j,\tau}}\right|^r \geq c_0 p_1^{-\delta r}\sum_{\tau=t^*+1}^{j}\left|\widehat{\lambda}_{k_1+1,\tau}\right|^r.
$$

Note also that, by Lemma C.3*(i)*, there are two numbers $0<a<b<1$ such that

$$
\lambda_{k_1+1,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau} E\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right)=c_0 p_1,
$$

for some $0 < c_0 < \infty$ and $t^* + am \le \tau \le t^* + m - bm$. On this interval, we study

$$\begin{aligned}
(\text{C.16}) \quad & \sum_{\tau=t^*+1}^{j} \left| \left|\widehat{\lambda}_{k_1+1,\tau}\right|^r - \left|\lambda_{k_1+1,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau} E\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right)\right|^r \right| \\
\le\ & r2^{r-1}\sum_{\tau=t^*+1}^{j}\left|\widehat{\lambda}_{k_1+1,\tau} - \lambda_{k_1+1,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau} E\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right)\right|^r \\
& + r2^{r-1}\sum_{\tau=t^*+1}^{j}\left|\lambda_{k_1+1,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau} E\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right)\right|^{r-1} \\
& \times \left|\widehat{\lambda}_{k_1+1,\tau} - \lambda_{k_1+1,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau} E\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right)\right|.
\end{aligned}$$

Note that

$$\begin{aligned}
& \left|\widehat{\lambda}_{k_1+1,\tau} - \lambda_{k_1+1,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau} E\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right)\right| \\
\le\ & \left|\lambda_{\max,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau}\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t'\mathbf{R}' - E\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right)\right)\right| + \left|\lambda_{\max,\tau}\left(\mathbf{M}_\tau^*\right)\right|,
\end{aligned}$$

where $\mathbf{M}_\tau^*$ is defined as

$$\mathbf{M}_\tau^* = \frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t' + \frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{E}_t' + \frac{1}{mp_2^2}\sum_{t=\tau+1-m}^{\tau}\mathbf{E}_t\widehat{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'.$$

We already know from the proof of Lemma C.5 that

$$(\text{C.17}) \qquad p_1^{-\delta r} T_m \sum_{t=\tau+1-m}^{\tau}\left|\lambda_{\max,\tau}\left(\mathbf{M}_\tau^*\right)\right|^r = o_{a.s.}(1).$$

Further, we have

$$\begin{aligned}
(\text{C.18}) \quad & \lambda_{\max,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau}\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\mathbf{C}\mathbf{F}_t'\mathbf{R}' - E\left(\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right)\right) \\
\le\ & \lambda_{\max,\tau}\left(\frac{1}{m}\mathbf{R}\sum_{t=\tau+1-m}^{\tau}\left(\mathbf{F}_t\mathbf{F}_t' - E\left(\mathbf{F}_t\mathbf{F}_t'\right)\right)\mathbf{R}'\right) \\
& + \lambda_{\max,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}} - \mathbf{C}\right)\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right) \\
& + \lambda_{\max,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\mathbf{C}\left(\widetilde{\mathbf{C}} - \mathbf{C}\right)'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right) \\
& + \lambda_{\max,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}} - \mathbf{C}\right)\left(\widetilde{\mathbf{C}} - \mathbf{C}\right)'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right).
\end{aligned}$$

Hence we have (assuming $\mathbf{F}_t$ scalar without loss of generality in the last passage)

$$
\begin{aligned}
& \sum_{\tau=t^*+am}^{t^*+m-bm} E\left|\lambda_{\max,\tau}\left(\frac{1}{m}\mathbf{R}\sum_{t=\tau+1-m}^{\tau}\left(\mathbf{F}_t\mathbf{F}_t' - E\left(\mathbf{F}_t\mathbf{F}_t'\right)\right)\mathbf{R}'\right)\right|^r \\
\leq\ & c_0 m^{-r} p_1^r \sum_{\tau=t^*+am}^{t^*+m-bm} E\left|\lambda_{\max,\tau}\left(\sum_{t=\tau+1-m}^{\tau}\left(\mathbf{F}_t\mathbf{F}_t' - E\left(\mathbf{F}_t\mathbf{F}_t'\right)\right)\right)\right|^r \\
\leq\ & c_0 m^{-r} p_1^r \sum_{\tau=t^*+am}^{t^*+m-bm} E\left|\sum_{t=\tau+1-m}^{\tau}\left(\mathbf{F}_t^2 - E\left(\mathbf{F}_t^2\right)\right)\right|^r \leq c_1 m^{1-r/2} p_1^r
\end{aligned}
$$

having used Assumption A.2*(ii)* in the second line, and Assumption A.1*(iii)* in the last one. Hence, Lemma C.1 ensures that

(C.19)
$$
\sum_{\tau=t^*+am}^{t^*+m-bm}\left|\lambda_{\max,\tau}\left(\frac{1}{m}\mathbf{R}\sum_{t=\tau+1-m}^{\tau}\left(\mathbf{F}_t\mathbf{F}_t' - E\left(\mathbf{F}_t\mathbf{F}_t'\right)\right)\mathbf{R}'\right)\right|^r = o_{a.s.}\left(m^{1-r/2} p_1^r \left(\ln m\right)^{1+1/r+\epsilon}\right).
$$

Similarly

$$
\begin{aligned}
& \sum_{\tau=t^*+am}^{t^*+m-bm} E\left|\lambda_{\max,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right|^r \\
\leq\ & c_0 p_2^{-r}\left\|\mathbf{R}\right\|_{\mathbf{F}}^{2r}\left\|\mathbf{C}\right\|_{\mathbf{F}}^{r}\left\|\widetilde{\mathbf{C}}-\mathbf{C}\right\|_{\mathbf{F}}^{r}\sum_{\tau=t^*+am}^{t^*+m-bm} E\left|\frac{1}{m}\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t^2\right|^r \\
\leq\ & c_0 p_1^r\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)^r\sum_{\tau=t^*+am}^{t^*+m-bm}\left(\frac{1}{m}\sum_{t=\tau+1-m}^{\tau}E\left|\mathbf{F}_t\right|^{2r}\right) \\
\leq\ & c_0 m\left(p_1\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)\right)^r,
\end{aligned}
$$

having used Lemma C.4 in the second line, and Assumption A.1*(i)* in the last line. Therefore, using Lemma C.1

(C.20)
$$
\begin{aligned}
& \sum_{\tau=t^*+am}^{t^*+m-bm} E\left|\lambda_{\max,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\mathbf{C}'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right|^r \\
=\ & o_{a.s.}\left(m\left(p_1\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)\right)^r\left(\ln m\right)^{1+1/r+\epsilon}\right).
\end{aligned}
$$

The same applies to the third term in (C.18) by symmetry. Finally

$$
\begin{aligned}
& \sum_{\tau=t^*+am}^{t^*+m-bm} E\left|\lambda_{\max,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right|^r \\
\leq\ & c_0 p_2^{-2r}\left\|\mathbf{R}\right\|_F^{2r}\left\|\mathbf{C}\right\|_F^{2r}\left\|\widetilde{\mathbf{C}}-\mathbf{C}\right\|_F^{2r}\sum_{\tau=t^*+am}^{t^*+m-bm} E\left|\frac{1}{m}\sum_{t=\tau+1-m}^{\tau}\mathbf{F}_t^2\right|^r \\
\leq\ & c_0 p_1^r\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)^{2r}\sum_{\tau=t^*+am}^{t^*+m-bm}\left(\frac{1}{m}\sum_{t=\tau+1-m}^{\tau} E\left|\mathbf{F}_t\right|^{2r}\right) \\
\leq\ & c_0 m\left(p_1^{1/2}\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)\right)^{2r},
\end{aligned}
$$

and, by Lemma C.1

$$
\begin{aligned}
& \sum_{\tau=t^*+am}^{t^*+m-bm} E\left|\lambda_{\max,\tau}\left(\frac{1}{p_2^2 m}\sum_{t=\tau+1-m}^{\tau}\mathbf{R}\mathbf{F}_t\mathbf{C}'\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)\left(\widetilde{\mathbf{C}}-\mathbf{C}\right)'\mathbf{C}\mathbf{F}_t'\mathbf{R}'\right)\right|^r \\
=\ & o_{a.s.}\left(m\left(p_1^{1/2}\left(\frac{1}{p_2}+\frac{1}{\sqrt{mp_1}}\right)\right)^{2r}(\ln m)^{1+1/r+\epsilon}\right),
\end{aligned}
\tag{C.21}
$$

so that this term is dominated by the previous one in (C.20). Therefore, combining (C.17) and (C.19)-(C.21), and using (C.16), the desired result follows. The proof of (C.15) follows from the same passages, upon noting that under (3.4), by (3.5), there is a number $0 < a < 1$ such that there exists a triplet of random variables $(p_{1,0}, p_{2,0}, T_0)$ and a positive constant c_0 such that, for $p_1 \geq p_{1,0}$, $p_2 \geq p_{2,0}$ and $T \geq T_0$

$$
\widehat{\lambda}_{k_1+1,\tau} \geq c_0 p_1 \text{ for } t^* + am \leq \tau \leq T_m.
$$

□

D. Proofs.

Proof of Theorem 1. We begin by showing (3.18), where $\eta < 1/2$. Define

$$S_\tau^* = \sum_{j=1}^{\tau} z_j,$$

and recall that $\{z_\tau, 1 \le \tau \le T_m\}$ is *i.i.d.* $N(0,1)$. Since $\psi_\tau \ge 0$ for all τ, it holds that

$$\max_{1\le\tau\le T_m} \frac{|S_\tau - S_\tau^*|}{\tau^\eta} = \max_{1\le\tau\le T_m} \frac{\sum_{j=1}^{\tau}\psi_j}{\tau^\eta} = T_m^{-\eta}\sum_{j=1}^{T_m}\psi_j = o_{P^*}(1), \tag{D.1}$$

due to Lemma C.5. By the KMT approximation (see Komlós et al., 1975, and Komlós et al., 1976), on a suitably enlarged probability space there exists a sequence of standard Wiener process $\{W_{T_m}(x), 0 < x < \infty\}$, such that

$$\max_{1\le\tau\le T_m} \frac{1}{\ln\tau}\left|S_\tau^* - W_{T_m}(\tau)\right| = O_P(1). \tag{D.2}$$

Then, using (D.2), it follows that

$$\begin{aligned} & T_m^{\eta-1/2}\max_{1\le\tau\le Tm} \tau^{-\eta}\left|S_\tau^* - W_{T_m}(\tau)\right| \\ \le\ & T_m^{\eta-1/2}\max_{1\le\tau\le Tm} \frac{1}{\ln\tau}\left|S_\tau^* - W_{T_m}(\tau)\right|\max_{1\le\tau\le Tm}\tau^{-\eta}\ln\tau \\ =\ & O_P\left(T_m^{\eta-1/2}\ln T_m\right) = o_P(1). \end{aligned} \tag{D.3}$$

Finally

$$\begin{aligned} & T_m^{\eta-1/2}\max_{1\le\tau\le Tm}\tau^{-\eta}\left|W_{T_m}(\tau)\right| \\ =\ & \max_{1\le\tau\le Tm}\left(\frac{\tau}{T_m}\right)^{-\eta}\left|W_{T_m}\left(\frac{\tau}{T_m}\right)\right| \\ \stackrel{D}{=}\ & \max_{1/T_m\le u\le 1}(u)^{-\eta}\left|W(u)\right| \\ \stackrel{a.s.}{\to}\ & \max_{0\le u\le 1}(u)^{-\eta}\left|W(u)\right|, \end{aligned} \tag{D.4}$$

as $T_m \to \infty$, having used the scale transformation of the Wiener process. Putting (D.1), (D.2), (D.3) and (D.4) together, (3.18) follows.

We now turn to showing (3.19). Note first that

$$\begin{aligned} & \left|\max_{1\le\tau\le T_m}\tau^{-1/2}|S_\tau| - \max_{1\le\tau\le T_m}\tau^{-1/2}|S_\tau^*|\right| \\ \le\ & \max_{1\le\tau\le T_m}\tau^{-1/2}\sum_{j=1}^{\tau}\psi_j \le T_m^{-1/2}\sum_{j=1}^{T_m}\psi_j = o_{a.s.}\left(T_m^{-1/2}\right), \end{aligned}$$

having used again Lemma C.5. Thus it holds that

$$\alpha_{Tm}\left|\max_{1\le\tau\le T_m}|S_\tau| - \max_{1\le\tau\le T_m}|S_\tau^*|\right| - \beta_{Tm} \stackrel{a.s.}{\to} -\infty; \tag{D.5}$$

this proves equation (2.3) in Shorack (1979). Hence

$$P^*\left(\alpha_{Tm}\max_{1\leq\tau\leq T_m}\tau^{-1/2}\left|S_\tau\right|\leq x+\beta_{Tm}\right)=P^*\left(\alpha_{Tm}\max_{1\leq\tau\leq T_m}\tau^{-1/2}\left|S_\tau^*\right|\leq x+\beta_{Tm}\right)+o_{P^*}\left(1\right);$$

recalling that S_τ^* is the partial sums process of a sequence of *i.i.d.* standard normals, the Darling-Erdős theorem (Darling and Erdős, 1956) yields the desired result.

Finally, we show (3.20). The proof is similar to the above, so we omit passages when this does not cause confusion. Note that

$$\begin{aligned}
& r_{T_m}^{\eta-1/2}\left|\max_{r_{T_m}\leq\tau\leq T_m}\tau^{-\eta}\left|S_\tau\right|-\max_{r_{T_m}\leq\tau\leq T_m}\tau^{-\eta}\left|S_\tau^*\right|\right| \\
\leq\ & r_{T_m}^{\eta-1/2}\max_{r_{T_m}\leq\tau\leq T_m}\tau^{-\eta}\sum_{j=1}^{\tau}\psi_j,
\end{aligned}$$

which - on account of Lemma C.5 - is bounded by $o_{a.s.}\left(r_{T_m}^{\eta-1/2}T_m^{-\eta}\right)$, whenever $\frac{1}{2}<\eta\leq 1$, and by $o_{a.s.}\left(r_{T_m}^{1/2}T_m^{-1}\right)$, when $\eta>1$, as $\min\left\{p_1,p_2,m\right\}\to\infty$. In both cases, the bounds drift to zero in light of (3.17). Also

$$\begin{aligned}
& r_{T_m}^{\eta-1/2}\max_{r_{T_m}\leq\tau\leq T_m}\tau^{-\eta}\left|S_\tau^*-W_{T_m}\left(\tau\right)\right| \\
\leq\ & r_{T_m}^{\eta-1/2}\max_{r_{T_m}\leq\tau\leq T_m}\tau^{-\eta}\frac{\left|S_\tau^*-W_{T_m}\left(\tau\right)\right|}{\ln\tau}\ln\tau \\
\leq\ & O_P\left(1\right)r_{T_m}^{\eta-1/2}\max_{r_{T_m}\leq\tau\leq T_m}\tau^{-\eta}\ln\tau \\
=\ & O_P\left(r_{T_m}^{-1/2}\ln r_{T_m}\right)=o_P\left(1\right),
\end{aligned}$$

where S_τ^* and $W_{T_m}\left(\tau\right)$ are defined above. Hence we need to study

$$r_{T_m}^{\eta-1/2}\max_{r_{T_m}\leq\tau\leq T_m}\tau^{-\eta}\left|W_{T_m}\left(\tau\right)\right|;$$

using the scale transformation, we obtain

$$r_{T_m}^{\eta-1/2}\max_{r_{T_m}\leq\tau\leq T_m}\tau^{-\eta}\left|W_{T_m}\left(\tau\right)\right|\overset{D}{=}\left(\frac{r_{T_m}}{T_m}\right)^{\eta-1/2}\max_{\frac{r_{T_m}}{T_m}\leq t\leq 1}t^{-\eta}\left|W\left(t\right)\right|;$$

setting $s=t\frac{T_m}{r_{T_m}}$, and using again the scale transformation

$$\left(\frac{r_{T_m}}{T_m}\right)^{\eta-1/2}\max_{\frac{r_{T_m}}{T_m}\leq t\leq 1}t^{-\eta}\left|W\left(t\right)\right|\overset{D}{=}\max_{1\leq s\leq\frac{T_m}{r_{T_m}}}s^{-\eta}\left|W\left(s\right)\right|\overset{a.s.}{\to}\sup_{1\leq s<\infty}\frac{\left|W\left(s\right)\right|}{s^\eta},$$

completing the proof. □

Proof of Theorem 2. We write

$$S_\tau=\sum_{j=1}^{\tau}z_j+\sum_{j=1}^{t^*}\psi_j+\sum_{j=t^*+1}^{\tau}\psi_j,$$

with the convention that $\sum_c^d = 0$ whenever $d < c$. We consider the alternative hypothesis (3.2); results under (3.4) follow from exactly the same logic.

We now turn to proving the theorem. Consider first the case $0 \leq \eta < \frac{1}{2}$, and note that, by the Law of the Iterated Logarithm

$$T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \tau^{-\eta} \sum_{j=1}^{\tau} z_j = O_{a.s.}\left(\sqrt{\ln\ln T_m}\right);$$

also, following the proof of Lemma C.5, it is easy to check that

$$T_m^{\eta-1/2}\left(\max_{1\leq\tau\leq T_m} \tau^{-\eta}\right) \sum_{j=1}^{t^*} \psi_j = o_{a.s.}\left(T_m^{\eta-1/2} m T_m^{-1}\right) = o_{a.s.}(1),$$

having used Assumption 4. Finally, by following the proof of Lemma C.6, there exists a triplet of random variables $(p_{1,0}, p_{2,0}, T_0)$ and a positive constant c_0 such that, for $p_1 \geq p_{1,0}$, $p_2 \geq p_{2,0}$ and $T \geq T_0$

$$\begin{aligned}
&\max_{1\leq\tau\leq T_m} \tau^{-\eta} \sum_{j=t^*+1}^{\tau} \psi_j \\
\geq\ & \max_{1\leq\tau\leq T_m} \tau^{-\eta} \sum_{j=am+t^*+1}^{\tau} \psi_j \\
\geq\ & c_0 p_1^{r(1-\delta)} \max_{am+t^*+1\leq\tau\leq m-bm+t^*} \tau^{-\eta}\left(\tau - t^*\right) \\
=\ & c_1 p_1^{r(1-\delta)} m^{1-\eta},
\end{aligned} \tag{D.6}$$

where the last result follows from standard algebra, and it is valid for all η. Now (3.23) follows putting everything together. When $\eta > 1/2$, essentially the same arguments yield

$$\begin{aligned}
r_{T_m}^{\eta-1/2} \max_{r_{T_m}\leq\tau\leq T_m} \tau^{-\eta} \sum_{j=1}^{\tau} z_j &= O_{a.s.}\left(\sqrt{\ln\ln T_m}\right), \\
r_{T_m}^{\eta-1/2} \max_{r_{T_m}\leq\tau\leq T_m} \tau^{-\eta} \sum_{j=1}^{t^*} \psi_j &= o_{a.s.}\left(r_{T_m}^{-1/2} m T_m^{-1}\right) = o_{a.s.}(1).
\end{aligned}$$

Equation (3.26) follows by putting these results and (D.6) together. Finally, when $\eta = 1/2$, it is easy to see by the same logic as above that

$$\begin{aligned}
\max_{1\leq\tau\leq T_m} \tau^{-1/2} \sum_{j=1}^{\tau} z_j &= O_{a.s.}\left(\sqrt{\ln\ln T_m}\right), \\
\max_{1\leq\tau\leq T_m} \tau^{-1/2} \sum_{j=1}^{t^*} \psi_j &= o_{a.s.}\left(m T_m^{-1}\right),
\end{aligned}$$

so that (3.24) follows from the same logic as above. Under (3.4), the proof follows essentially by the same arguments.

□

Proof of Theorem 3. We will use the short-hand notation

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

Note that

$$P^*\left(\frac{Z_{Tm}-b_{Tm}}{a_{Tm}}\leq v\right) = P^*\left(Z_{Tm}\leq a_{Tm}v+b_{Tm}\right),$$

and consider the sequence y_τ defined in (3.11), viz.

$$y_\tau = z_\tau + \psi_\tau.$$

By construction, z_τ is independent across τ and independent of the sample $\{\mathbf{X}_t, 1\leq t\leq T\}$. Hence, letting $\Phi(\cdot)$ be the distribution of the standard normal and $\varphi(\cdot)$ its density, we can write

$$\begin{aligned} P^*\left(Z_{Tm}\leq a_{Tm}v+b_{Tm}\right) &= \prod_{\tau=1}^{T_m} P^*\left(y_\tau\leq a_{Tm}v+b_{Tm}\right) \\ &= \prod_{\tau=1}^{T_m} P^*\left(z_\tau\leq a_{Tm}v+b_{Tm}-\psi_\tau\right) \\ &= \exp\left(\sum_{\tau=1}^{T_m}\ln\Phi\left(a_{Tm}v+b_{Tm}-\psi_\tau\right)\right). \end{aligned} \tag{D.7}$$

Note that

$$\ln\Phi\left(a_{Tm}v+b_{Tm}-\psi_\tau\right) = \ln\Phi\left(a_{Tm}v+b_{Tm}\right) + \ln\frac{\Phi\left(a_{Tm}v+b_{Tm}-\psi_\tau\right)}{\Phi\left(a_{Tm}v+b_{Tm}\right)}. \tag{D.8}$$

Using Taylor's expansion, for some $c_0>0$ it holds that

$$\begin{aligned} \ln\Phi\left(a_{Tm}v+b_{Tm}-\psi_\tau\right) &= \ln\Phi\left(a_{Tm}v+b_{Tm}\right) + \ln\frac{\Phi\left(a_{Tm}v+b_{Tm}-\psi_\tau\right)}{\Phi\left(a_{Tm}v+b_{Tm}\right)} \\ &= \ln\Phi\left(a_{Tm}v+b_{Tm}\right) + \ln\left[1-c_0\psi_\tau\frac{\varphi\left(a_{Tm}v+b_{Tm}\right)}{\Phi\left(a_{Tm}v+b_{Tm}\right)}\right]. \end{aligned} \tag{D.9}$$

Therefore, putting together (D.7)-(D.9)

$$\begin{aligned} &P^*\left(Z_{Tm}\leq a_{Tm}v+b_{Tm}\right) \\ =\ &\exp\left(\sum_{\tau=1}^{T_m}\ln\Phi\left(a_{Tm}v+b_{Tm}-\psi_\tau\right)\right) \\ =\ &\exp\left(\sum_{\tau=1}^{T_m}\ln\Phi\left(a_{Tm}v+b_{Tm}\right)\right)\exp\left(\sum_{\tau=1}^{T_m}\ln\left[1-c_0\psi_\tau\frac{\varphi\left(a_{Tm}v+b_{Tm}\right)}{\Phi\left(a_{Tm}v+b_{Tm}\right)}\right]\right) \\ =\ &\left[\Phi\left(a_{Tm}v+b_{Tm}\right)\right]^{T_m}\exp\left(\sum_{\tau=1}^{T_m}\ln\left[1-c_1\psi_\tau\right]\right). \end{aligned} \tag{D.10}$$

Consider the elementary inequality $\ln(1-c_1\psi_\tau)\geq -c_2\psi_\tau$, for some $c_2>0$; thus, from (D.10)

$$1\geq\exp\left(\sum_{\tau=1}^{T_m}\ln\left[1-c_1\psi_\tau\right]\right)\geq\exp\left(\sum_{\tau=1}^{T_m}-c_2\psi_\tau\right).$$

By Lemma C.5, we know that $-c_2 \sum_{\tau=1}^{T_m} \psi_\tau = o_{a.s.}(1)$, so that the final result obtains by dominated convergence.

Under both alternatives (3.2) and (3.4), it holds that there is a constant c_0 and a random variable m_0 such that, for $m \geq m_0$, we have

$$\widehat{\lambda}_{k_1+1,\tau} \geq c_0 p_1, \tag{D.11}$$

for at least one $t^* \leq \widetilde{\tau}$. Thus we have

$$P^*\left(Z_{Tm} \leq a_{Tm} v + b_{Tm}\right) = \prod_{\tau=1}^{T_m} P^*\left(y_\tau \leq a_{Tm} v + b_{Tm}\right) \leq P^*\left(y_{\widetilde{\tau}} \leq a_{Tm} v + b_{Tm}\right).$$

Note that

$$P^*\left(y_{\widetilde{\tau}} \leq a_{Tm} v + b_{Tm}\right) = P^*\left(z_{\widetilde{\tau}} \leq a_{Tm} v + b_{Tm} - \psi_{\widetilde{\tau}}\right) = \Phi\left(a_{Tm} v + b_{Tm} - \psi_{\widetilde{\tau}}\right).$$

Using equation (5) in Borjesson and Sundberg (1979)

$$\Phi\left(a_{Tm} v + b_{Tm} - \psi_\tau\right) \leq \frac{\exp\left(-\frac{1}{2}\left(a_{Tm} v + b_{Tm} - \psi_{\widetilde{\tau}}\right)^2\right)}{\sqrt{2\pi}\left|a_{Tm} v + b_{Tm} - \psi_{\widetilde{\tau}}\right|}.$$

By (D.11), it follows that

$$P^*\left(\omega : \lim_{p_1, p_2, m \to \infty} a_{Tm} v + b_{Tm} - \psi_{\widetilde{\tau}} = -\infty\right) = 1,$$

for all $-\infty < v < \infty$, because a_{Tm} and b_{Tm} are both $O\left(\sqrt{\ln T_m}\right)$. Therefore it follows that, as $\min\{p_1, p_2, m\} \to \infty$, $P^*\left(z_\tau \leq a_{Tm} v + b_{Tm} - \psi_{\widetilde{\tau}}\right) = 0$ a.s. conditionally on the sample. Thus, by dominated convergence, as $\min\{p_1, p_2, m\} \to \infty$ it holds that

$$P^*\left(Z_{\widetilde{\tau}} \leq a_{Tm} v + b_{Tm}\right) = 0, \quad a.s.$$

conditionally on the sample. This implies the desired result. □

Proof of Proposition A.1. The proof uses similar arguments as above and as the proof of Theorem A.1, and therefore we only report the main arguments. We begin by showing that, under H_0, it holds that

$$\sum_{\tau=1}^{T_m} \widetilde{\psi}_\tau = o_{a.s.}(1). \tag{D.12}$$

Indeed, using Lemma C.2 there exists a constant $c_0 < \infty$ and a triplet of random variables $(p_{1,0}, p_{2,0}, m_0)$ such that, for all $p_1 \geq p_{1,0}$, $p_2 \geq p_{2,0}$ and $m \geq m_0$

$$\sum_{\tau=1}^{T_m} \widetilde{\psi}_\tau \leq c_0 \sum_{\tau=1}^{T_m} \left|p_1^{-\delta} \widehat{\lambda}_{k_1,\tau}\right|^{-r}.$$

We know from the proof of Lemma C.4 that

$$\sum_{\tau=1}^{T_m} \left|p_1^{-1} \widehat{\lambda}_{k_1,\tau} - p_1^{-1} \lambda_{k_1,\tau}\left(\frac{1}{m} \sum_{t=1}^{m} \mathbf{R} \mathbf{F}_t \mathbf{F}_t' \mathbf{R}'\right)\right|^r = o_{a.s.}(1),$$

Moreover, it is easy to see, following the same arguments as in the proofs above, that

$$E\left|\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}\mathbf{F}_t\mathbf{F}_t'\mathbf{R}'\right)-\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}E\left(\mathbf{F}_t\mathbf{F}_t'\right)\mathbf{R}'\right)\right|^r\leq c_0p_1^rm^{-r/2}.$$

Lemma 1 states that

$$\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}E\left(\mathbf{F}_t\mathbf{F}_t'\right)\mathbf{R}'\right)\geq c_0p_1.$$

Hence, combining the three equations above, we have

$$\begin{aligned}
&\sum_{\tau=1}^{T_m}\widetilde{\psi}_\tau\\
\leq\ & c_0\sum_{\tau=1}^{T_m}\left|p_1^{-\delta}\widehat{\lambda}_{k_1,\tau}\right|^{-r}\\
=\ & c_0p_1^{-\delta r}\sum_{\tau=1}^{T_m}\left|\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}E\left(\mathbf{F}_t\mathbf{F}_t'\right)\mathbf{R}'\right)+\widehat{\lambda}_{k_1,\tau}-\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}E\left(\mathbf{F}_t\mathbf{F}_t'\right)\mathbf{R}'\right)\right|^{-r}\\
=\ & c_0p_1^{\delta r}\left|\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}E\left(\mathbf{F}_t\mathbf{F}_t'\right)\mathbf{R}'\right)\right|^{-r}\sum_{\tau=1}^{T_m}\left|1+\frac{\widehat{\lambda}_{k_1,\tau}-\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}E\left(\mathbf{F}_t\mathbf{F}_t'\right)\mathbf{R}'\right)}{\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}E\left(\mathbf{F}_t\mathbf{F}_t'\right)\mathbf{R}'\right)}\right|^{-r}\\
\leq\ & c_1p_1^{(1-\delta)r}\sum_{\tau=1}^{T_m}\left|1+\frac{\widehat{\lambda}_{k_1,\tau}-\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}E\left(\mathbf{F}_t\mathbf{F}_t'\right)\mathbf{R}'\right)}{\lambda_{k_1,\tau}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbf{R}E\left(\mathbf{F}_t\mathbf{F}_t'\right)\mathbf{R}'\right)}\right|^{-r}\\
=\ & O_{a.s.}\left(\frac{m}{p_1^{r(1-\delta)}}\right)=o_{a.s.}(1),
\end{aligned}$$

having used Taylor's expansion in the penultimate passage, and (B.4) in the last one. Hereafter, the proof is the same as that of Theorems 1 and 3. □

Proof of Theorem A.1. Some arguments in the proof are similar to the arguments used in the proofs above, and we therefore omit them when possible. We begin by considering the behaviour under the null, and show that

$$m^{a(r/2-1)}\sum_{\tau=1}^{T_m}\left|1-\frac{\widehat{\lambda}_{k_1,\tau}}{\widehat{\lambda}_{k_1,0}}\right|^r=o_{a.s.}\left(m^{a(r/2-1)+1}\left(\frac{1}{\sqrt{m}}+\frac{1}{p_1}l_{p_1,p_2,m}+\frac{1}{p_2}\right)^r\right)=o_{a.s.}(1),\tag{D.13}$$

under (B.7). It holds that

$$\sum_{\tau=1}^{T_m}\left|1-\frac{\widehat{\lambda}_{k_1,\tau}}{\widehat{\lambda}_{k_1,0}}\right|^r=\sum_{\tau=1}^{T_m}\left|\frac{\widehat{\lambda}_{k_1,0}-\widehat{\lambda}_{k_1,\tau}}{\widehat{\lambda}_{k_1,0}}\right|^r=\widehat{\lambda}_{k_1,0}^{-r}\sum_{\tau=1}^{T_m}\left|\widehat{\lambda}_{k_1,0}-\widehat{\lambda}_{k_1,\tau}\right|^r,$$

and we already know that $\widehat{\lambda}_{k_1,0}\geq c_0p_1$ a.s. by virtue of Lemma 1, whence we write

$$\sum_{\tau=1}^{T_m}\left|1-\frac{\widehat{\lambda}_{k_1,\tau}}{\widehat{\lambda}_{k_1,0}}\right|^r\leq c_0p_1^{-r}\sum_{\tau=1}^{T_m}\left|\widehat{\lambda}_{k_1,0}-\widehat{\lambda}_{k_1,\tau}\right|^r,$$

and focus on $\sum_{\tau=1}^{T_m} \left| \widehat{\lambda}_{k_1,0} - \widehat{\lambda}_{k_1,\tau} \right|^r$. We use the estimate

$$\sum_{\tau=1}^{T_m} \left| \widehat{\lambda}_{k_1,0} - \widehat{\lambda}_{k_1,\tau} \right|^r \leq \sum_{\tau=1}^{T_m} \left\| \widehat{\mathbf{M}}_{1,0} - \widehat{\mathbf{M}}_{1,\tau} \right\|_F^r,$$

where, as far as the notation is concerned, $\widehat{\mathbf{M}}_{1,0}$ is $\widehat{\mathbf{M}}_1$ computed using data $1 \leq t \leq m$, and $\widehat{\mathbf{M}}_{1,\tau}$ is computed using data in the monitoring horizon $\tau+1 \leq t \leq m+\tau$. We have

$$\begin{aligned}
& \left\| \widehat{\mathbf{M}}_{1,0} - \widehat{\mathbf{M}}_{1,\tau} \right\|_F^r \\
\leq & \left\| \frac{1}{p_2^2 m} \sum_{t=1}^{m} \mathbf{R}\mathbf{F}_t \mathbf{C}' \widetilde{\mathbf{C}} \widetilde{\mathbf{C}}' \mathbf{C} \mathbf{F}_t' \mathbf{R}' - \frac{1}{p_2^2 m} \sum_{t=\tau+1}^{m+\tau} \mathbf{R}\mathbf{F}_t \mathbf{C}' \widetilde{\mathbf{C}} \widetilde{\mathbf{C}}' \mathbf{C} \mathbf{F}_t' \mathbf{R}' \right\|_F^r \\
& + \left\| \rho\left(\widehat{\mathbf{M}}_{1,0} \right) \right\|_F^r + \left\| \rho\left(\widehat{\mathbf{M}}_{1,\tau} \right) \right\|_F^r
\end{aligned}$$

where $\rho\left(\widehat{\mathbf{M}}_{1,0}\right)$ are the remainder terms in $\widehat{\mathbf{M}}_{1,0}$, and $\rho\left(\widehat{\mathbf{M}}_{1,\tau}\right)$ is defined similarly. Following the proof of Lemma C.5, it is easy to obtain that

$$\widehat{\lambda}_{k_1,0}^{-r} \sum_{\tau=1}^{T_m} \left\| \rho\left(\widehat{\mathbf{M}}_{1,\tau} \right) \right\|_{\mathbf{F}}^r = o_{a.s.}\left(T_m \left(p_1^{-1} l_{p_1,p_2,m} \right)^r \right),$$

and the same for $\widehat{\lambda}_{k_1,0}^{-r} \sum_{\tau=1}^{T_m} \left\| r\left(\widehat{\mathbf{M}}_{1,0} \right) \right\|_{\mathbf{F}}^r$. By (B.7), these terms, pre-multiplied by $m^{a(r/2-1)}$, are both $o_{a.s.}(1)$. Similarly, the expression

$$\widehat{\lambda}_{k_1,0}^{-r} \sum_{\tau=1}^{T_m} \left\| \frac{1}{p_2^2 m} \sum_{t=1}^{m} \mathbf{R}\mathbf{F}_t \mathbf{C}' \widetilde{\mathbf{C}} \widetilde{\mathbf{C}}' \mathbf{C} \mathbf{F}_t' \mathbf{R}' - \frac{1}{m} \sum_{t=1}^{m} \mathbf{R}\mathbf{F}_t \mathbf{F}_t' \mathbf{R}' \right\|_F$$

contains terms like (or of higher order than)

$$\begin{aligned}
& \widehat{\lambda}_{k_1,0}^{-r} \sum_{\tau=1}^{T_m} \left\| \frac{1}{p_2^2 m} \sum_{t=1}^{m} \mathbf{R}\mathbf{F}_t \mathbf{C}' \left(\widetilde{\mathbf{C}} - \mathbf{C} \right) \widetilde{\mathbf{C}}' \mathbf{C} \mathbf{F}_t' \mathbf{R}' \right\|_F^r \\
\leq & \widehat{\lambda}_{k_1,0}^{-r} \|\mathbf{R}\|_F^{2r} \left(\left\| p_2^{-1/2} \mathbf{C} \right\|_F^r \left\| p_2^{-1/2} \mathbf{C} \right\|_F^r \left\| p_2^{-1/2} \widetilde{\mathbf{C}} \right\|_F^r \right) \left(\left\| p_2^{-1/2} \left(\widetilde{\mathbf{C}} - \mathbf{C} \right) \right\|_F^r \right) \sum_{\tau=1}^{T_m} \left\| \frac{1}{m} \sum_{t=1}^{m} \mathbf{F}_t \mathbf{F}_t' \right\|_F^r \\
\leq & c_0 \left(\frac{1}{p_2} + \frac{1}{\sqrt{m p_1}} \right)^r T_m,
\end{aligned}$$

which, after multiplynig by $m^{a(r/2-1)}$, are $o_{a.s.}(1)$ in the light of (B.7) and given that $r > 2$. Finally we bound

$$\begin{aligned}
& \widehat{\lambda}_{k_1,0}^{-r} \sum_{\tau=1}^{T_m} \left\| \frac{1}{m} \sum_{t=1}^{m} \left(\mathbf{R}\mathbf{F}_t \mathbf{F}_t' \mathbf{R}' - \mathbf{R}\mathbf{\Sigma}_1 \mathbf{R}' \right) - \frac{1}{m} \sum_{t=\tau+1}^{m+\tau} \left(\mathbf{R}\mathbf{F}_t \mathbf{F}_t' \mathbf{R}' - \mathbf{R}\mathbf{\Sigma}_1 \mathbf{R}' \right) \right\|_F^r \\
\leq & \widehat{\lambda}_{k_1,0}^{-r} \sum_{\tau=1}^{T_m} \left\| \frac{1}{m} \sum_{t=1}^{m} \left(\mathbf{R}\mathbf{F}_t \mathbf{F}_t' \mathbf{R}' - \mathbf{R}\mathbf{\Sigma}_1 \mathbf{R}' \right) \right\|_F^r + \widehat{\lambda}_{k_1,0}^{-r} \sum_{\tau=1}^{T_m} \left\| \frac{1}{m} \sum_{t=\tau+1}^{m+\tau} \left(\mathbf{R}\mathbf{F}_t \mathbf{F}_t' \mathbf{R}' - \mathbf{R}\mathbf{\Sigma}_1 \mathbf{R}' \right) \right\|_F^r
\end{aligned}$$

where $\mathbf{\Sigma}_1$ is defined in Assumption A.1. By standard arguments

$$\begin{aligned}
& m^{a(r/2-1)}\widehat{\lambda}_{k_1,0}^{-r}\sum_{\tau=1}^{T_m}\left\|\frac{1}{m}\sum_{t=\tau+1}^{m+\tau}\left(\mathbf{R}\mathbf{F}_t\mathbf{F}_t'\mathbf{R}'-\mathbf{R}\mathbf{\Sigma}_1\mathbf{R}'\right)\right\|_F^r \\
\leq \quad & c_0 m^{a(r/2-1)}p_1^{-r}\left\|\mathbf{R}\right\|_F^{2r}\sum_{\tau=1}^{T_m}\left\|\frac{1}{m}\sum_{t=\tau+1}^{m+\tau}\left(\mathbf{F}_t\mathbf{F}_t'-\mathbf{\Sigma}_1\right)\right\|_F^r \\
= \quad & o_{a.s.}\left(m^{a(r/2-1)}\frac{T_m}{m^{r/2}}\left(\ln m\right)^{1+\epsilon}\right)=o_{a.s.}\left(1\right).
\end{aligned}$$

Putting all together, (D.13) follows. Under the alternative, it immediately follows that there is a noncentrality term given by

$$m^{a(r/2-1)}\sum_{\tau=t^*+1}^{T_m}\left|\frac{\tau-t^*}{m}\right|^r. \tag{D.14}$$

Henceforth, the proof is the same as that of Theorems 1-3. □

Proof of Theorem A.2. The proof is immediate upon noting that, under the null, following the proof of Theorem 1 and recalling (3.12)

$$\max_{a_m\leq\tau\leq T_m}\frac{\left|S_\tau-W_{T_m}\left(\tau\right)\right|}{\tau^{1/2}}=\max_{a_m\leq\tau\leq T_m}\frac{\left|\sum_{j=1}^{\tau}z_\tau-W_{T_m}\left(\tau\right)\right|}{\tau^{1/2}}+o_{a.s.}\left(T_m^{-1/2}\right),$$

where the remainder is dominated by $o\left(a_m^{-1/2}\ln a_m\right)$, by the definition of a_m. Using the KMT approximation (Komlós et al., 1975, and Komlós et al., 1976) we have

$$\begin{aligned}
& \max_{a_m\leq\tau\leq T_m}\frac{\left|\sum_{j=1}^{\tau}z_\tau-W_{T_m}\left(\tau\right)\right|}{\tau^{1/2}} \\
= \quad & \max_{a_m\leq\tau\leq T_m}\frac{\ln\tau}{\ln\tau}\frac{\left|\sum_{j=1}^{\tau}z_\tau-W_{T_m}\left(\tau\right)\right|}{\tau^{1/2}} \\
\leq \quad & \max_{a_m\leq\tau\leq T_m}\frac{1}{\ln\tau}\left|\sum_{j=1}^{\tau}z_\tau-W_{T_m}\left(\tau\right)\right|\max_{a_m\leq\tau\leq T_m}\frac{\ln\tau}{\tau^{1/2}} \\
= \quad & O_P\left(1\right)\max_{a_m\leq\tau\leq T_m}\frac{\ln\tau}{\tau^{1/2}}=O_P\left(\frac{\ln a_m}{a_m^{1/2}}\right).
\end{aligned} \tag{D.15}$$

Putting all together, the desired result follows. We note further that, under the assumptions of Theorem A.3, our results still hold, but (D.15) becomes

$$\begin{aligned}
& \max_{a_m\leq\tau\leq T_m}\frac{\left|\sum_{j=1}^{\tau}z_\tau-W_{T_m}\left(\tau\right)\right|}{\tau^{1/2}} \\
= \quad & \max_{a_m\leq\tau\leq T_m}\frac{\tau^{1/\gamma}}{\tau^{1/\gamma}}\frac{\left|\sum_{j=1}^{\tau}z_\tau-W_{T_m}\left(\tau\right)\right|}{\tau^{1/2}} \\
= \quad & O_P\left(1\right)\max_{a_m\leq\tau\leq T_m}\tau^{1/\gamma-1/2}=O_P\left(a_m^{1/\gamma-1/2}\right),
\end{aligned} \tag{D.16}$$

and therefore the approximation rate is worse than if the z_τs were Gaussian (or, more generally, admitted moments of order higher than γ). $\square$

PROOF OF THEOREM A.3. The proof is a standard consequence of the the KMT approximation (Komlós et al., 1975, and Komlós et al., 1976). Indeed, recall the notation $S_\tau^* = \sum_{j=1}^{\tau} z_j$. By the KMT approximation (Komlós et al., 1975, and Komlós et al., 1976), it holds that- on a suitably enlarged probability space - there exists a sequence of standard Wiener process $\{W_{T_m}(x), 0 < x < \infty\}$, such that

$$\max_{1\leq\tau\leq T_m} \frac{1}{\tau^{1/\gamma}} \left|S_\tau^* - W_{T_m}(\tau)\right| = O_P(1).$$

Whenever $\eta < 1/2$, this immediately entails that

$$\begin{aligned}
& T_m^{\eta-1/2} \max_{1\leq\tau\leq Tm} \tau^{-\eta} \left|S_\tau^* - W_{T_m}(\tau)\right| \\
\leq\ & T_m^{\eta-1/2} \max_{1\leq\tau\leq Tm} \frac{1}{\tau^{1/\gamma}} \left|S_\tau^* - W_{T_m}(\tau)\right| \max_{1\leq\tau\leq Tm} \tau^{1/\gamma-\eta} \\
=\ & o_P\left(T_m^{1/\gamma-1/2}\right) = o_P(1).
\end{aligned}$$

Equation (3.18) of Theorem 1 now follows immediately. Equation (3.20) of Theorem 1 can be shown in a similar way upon noting that

$$\begin{aligned}
& r_{T_m}^{\eta-1/2} \max_{r_{T_m}\leq\tau\leq T_m} \tau^{-\eta} \left|S_\tau^* - W_{T_m}(\tau)\right| \\
\leq\ & r_{T_m}^{\eta-1/2} \max_{r_{T_m}\leq\tau\leq T_m} \frac{\left|S_\tau^* - W_{T_m}(\tau)\right|}{\tau^{1/\gamma}} \max_{r_{T_m}\leq\tau\leq T_m} \tau^{1/\gamma}\tau^{-\eta} \\
=\ & O_P(1)\, r_{T_m}^{\eta-1/2} \max_{r_{T_m}\leq\tau\leq T_m} \tau^{1/\gamma-\eta} \\
=\ & O_P\left(r_{T_m}^{1/\gamma-1/2}\right) = o_P(1).
\end{aligned}$$

We now show that (3.19) also holds. We have

$$\begin{aligned}
& \max_{1\leq\tau\leq T_m} \frac{\left|S_\tau^* - W_{T_m}(\tau)\right|}{\tau^{1/2}} \\
\leq\ & \max_{1\leq\tau\leq T_m} \frac{\left|S_\tau^* - W_{T_m}(\tau)\right|}{\tau^{1/\gamma}} \max_{1\leq\tau\leq T_m} \tau^{1/\gamma-1/2} = O_P(1),
\end{aligned}$$

and therefore

$$a_{T_m} \max_{1\leq\tau\leq T_m} \frac{\left|S_\tau^* - W_{T_m}(\tau)\right|}{\tau^{1/2}} - b_{T_m} \to -\infty,$$

whence the desired result follows from the Darling-Erdős theorem (Darling and Erdős, 1956). Finally, the results in Theorem 2 can be shown by repeating, *verbatim*, the proofs of that theorem. $\square$

PROOF OF COROLLARY A.1. We prove the theorem considering, as a leading example, equation (3.18) - i.e., studying the case

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

where $\eta < 1/2$, with (k_1, k_2) replaced by $\left(\widetilde{k}_1, \widetilde{k}_2\right)$. All the other cases covered by the corollary can be studied with exactly the same logic. We note that, when using k_1 and k_2, (3.18) implies that

$$P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S(\tau)|}{\tau^\eta} \leq c_{\alpha,\eta}\right) = 1-\alpha,$$

as $T_m \to \infty$, for almost all realisations of $\{\mathbf{X}_t, 1 \leq t \leq T\}$. Recall that He et al. (2023) show that, as $\min\{p_1, p_2, m\} \to \infty$

$$\widetilde{k}_1 = k_1 + o_{P^*}(1), \quad \text{and} \quad \widetilde{k}_2 = k_2 + o_{P^*}(1), \tag{D.17}$$

for almost all realisations of $\{\mathbf{X}_t, 1 \leq t \leq T\}$; based on the above, it is easy to see that

$$P^*\left(\left(\widetilde{k}_1, \widetilde{k}_2\right) \neq (k_1, k_2)\right) = o_{P^*}(1), \tag{D.18}$$

for almost all realisations of $\{\mathbf{X}_t, 1 \leq t \leq T\}$.

We are now ready to prove our main result. It holds that

$$\begin{aligned}
& P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta} \leq c_{\alpha,\eta}\right) \\
= \ & P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta} \leq c_{\alpha,\eta} | \widetilde{k}_1 = k_1, \widetilde{k}_2 = k_2\right) P^*\left(\widetilde{k}_1 = k_1 | \widetilde{k}_2 = k_2\right) P^*\left(\widetilde{k}_2 = k_2\right) \\
& + P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta} \leq c_{\alpha,\eta} | \widetilde{k}_1 = k_1, \widetilde{k}_2 = k_2\right) P^*\left(\widetilde{k}_1 = k_1 | \widetilde{k}_2 \neq k_2\right) P^*\left(\widetilde{k}_2 \neq k_2\right) \\
& + P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta} \leq c_{\alpha,\eta} | \left(\widetilde{k}_1, \widetilde{k}_2\right) \neq (k_1, k_2)\right) P^*\left(\left(\widetilde{k}_1, \widetilde{k}_2\right) \neq (k_1, k_2)\right).
\end{aligned}$$

We now have that

$$\begin{aligned}
& P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta} \leq c_{\alpha,\eta} | \widetilde{k}_1 = k_1, \widetilde{k}_2 = k_2\right) P^*\left(\widetilde{k}_1 = k_1 | \widetilde{k}_2 \neq k_2\right) P^*\left(\widetilde{k}_2 \neq k_2\right) \\
\leq \ & P^*\left(\widetilde{k}_2 \neq k_2\right) = o_{P^*}(1),
\end{aligned}$$

by (D.17). Further

$$\begin{aligned}
& P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S(\tau)|}{\tau^\eta} \leq c_{\alpha,\eta} | \left(\widehat{k}_1, \widehat{k}_2\right) \neq (k_1, k_2)\right) P^*\left(\left(\widehat{k}_1, \widehat{k}_2\right) \neq (k_1, k_2)\right) \\
\leq \ & P^*\left(\left(\widehat{k}_1, \widehat{k}_2\right) \neq (k_1, k_2)\right) = o_{P^*}(1),
\end{aligned}$$

having used (D.18). Hence we have

$$\begin{aligned}
& P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta} \leq c_{\alpha,\eta}\right) \\
= \ & P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta} \leq c_{\alpha,\eta} | \widetilde{k}_1 = k_1, \widetilde{k}_2 = k_2\right) P^*\left(\widetilde{k}_1 = k_1 | \widetilde{k}_2 = k_2\right) P^*\left(\widetilde{k}_2 = k_2\right) + o_{P^*}(1) \\
= \ & P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta} \leq c_{\alpha,\eta} | \widetilde{k}_1 = k_1, \widetilde{k}_2 = k_2\right) P^*\left(\left(\widehat{k}_1, \widehat{k}_2\right) = (k_1, k_2)\right) + o_{P^*}(1) \\
= \ & P^*\left(T_m^{\eta-1/2} \max_{1\leq\tau\leq T_m} \frac{|S_\tau|}{\tau^\eta} \leq c_{\alpha,\eta} | \widetilde{k}_1 = k_1, \widetilde{k}_2 = k_2\right) + o_{P^*}(1),
\end{aligned}$$

which proves the theorem. $\square$

References.

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

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

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

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

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

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

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

Berkes, I., Hörmann, S., Schauer, J., 2011. Split invariance principles for stationary processes. Annals of Probability 39, 2441–2473.

Borjesson, P., Sundberg, C.E., 1979. Simple approximations of the error function q (x) for communications applications. IEEE Transactions on Communications 27, 639–643.

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

Chen, W., Lam, C., 2022. Rank and factor loadings estimation in time series tensor factor model by pre-averaging. arXiv preprint arXiv:2208.04012 .

Csörgo, M., Révész, P., 2014. Strong approximations in probability and statistics. Academic press.

Darling, D.A., Erdős, P., 1956. A limit theorem for the maximum of normalized sums of independent random variables. Duke Math. J 23, 143–155.

Gomes, M.I., Haan, L.D., 1999. Approximation by penultimate extreme value distributions. Extremes 2, 71–85.

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

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

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

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

Liu, W., Lin, Z., 2009. Strong approximation for a class of stationary processes. Stochastic Processes and their Applications 119, 249–280.

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

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

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

Miller, C., 1994. Three theorems on ρ^* random fields. Journal of Theoretical Probability 7, 867–882.

Rio, E., 1995. A maximal inequality and dependent Marcinkiewicz-Zygmund strong laws. Annals of Probability 23, 918–937.

Shao, Q.M., 1995. Maximal inequalities for partial sums of ρ-mixing sequences. Annals of Probability , 948–965.

Shorack, G.R., 1979. Extension of the darling and erdos theorem on the maximum of normalized sums. Annals of Probability 7, 1092–1096.

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

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

Uematsu, Y., Yamagata, T., 2021. Inference in sparsity-induced weak factor models. Journal of Business & Economic Statistics , 1–14.

Vostrikova, L.Y., 1982. Detection of a "disorder" in a wiener process. Theory of Probability & Its Applications 26, 356–362.

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

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

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

Zhang, L., 1998. Rosenthal type inequalities for B-valued strong mixing random fields and their applications. Science in China Series A: Mathematics 41, 736–745.

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

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

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

Long Yu
Department of Statistics
and Applied Probability,
National University of Singapore,
117546, Singapore
E-mail: stayl@nus.edu.sg