Journal of Multivariate Analysis 169 (2019) 138–165

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Asymptotics for empirical eigenvalue processes in high-dimensional linear factor models

Lajos Horváth [a], Gregory Rice [b,*]

[a] Department of Mathematics, University of Utah, Salt Lake City, UT 84112–0090, USA

[b] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

ARTICLE INFO

Article history:
Received 23 July 2017
Available online 7 July 2018

Keywords:
Change point analysis
Linear factor models
Principal component analysis

ABSTRACT

When vector-valued observations are of high dimension N relative to the sample size T, it is common to employ a linear factor model in order to estimate the underlying covariance structure or to further understand the relationship between coordinates. Asymptotic analyses of such models often consider the case in which both N and T tend jointly to infinity. Within this framework, we derive weak convergence results for processes of partial sample estimates of the largest eigenvalues of the sample covariance matrix. It is shown that if the effect of the factors is sufficiently strong, then the processes associated with the largest eigenvalues have Gaussian limits under general conditions on the divergence rates of N and T, and the underlying observations. If the common factors are "weak", then N must grow much more slowly in relation to T in order for the largest eigenvalue processes to have a Gaussian limit. We apply these results to develop general tests for structural stability of linear factor models that are based on measuring the fluctuations in the largest eigenvalues throughout the sample, which we investigate further by means of a Monte Carlo simulation study and an application to US treasury yield curve data.

1. Introduction

The largest eigenvalues and corresponding eigenvectors of the sample covariance matrix of vector-valued observations are frequently used as a simplified summary of their variability and dependence structure, especially if the dimension N of the observations is large. This general practice is commonly referred to as principal component analysis (PCA), and within the last two decades there has been a surge of research in both the statistics and econometrics communities aiming to understand the asymptotic properties of PCA when N is large in relation to the sample size T; for a brief summary of this literature we refer to [4,12,21,32]. These asymptotic analyses are often considered within the framework of linear factor models, which basically posit that the dependence between coordinates may be explained by a common linear dependence on a small number of random "factors". Under such models and as $N \to \infty$, the largest eigenvalues of the covariance matrix diverge assuming that the factor loadings do not shrink too much as N increases. This observation seems to have given rise to more general spiked covariance models, and asymptotics for the largest sample eigenvalues under such models are considered in [33,41]. An important distinction in each of these works is the relative rate at which N may increase with respect to T in order for the asymptotics to hold.

By and large most papers in this direction assume the vector-valued observations form a simple random sample, although in many arenas of application, such as finance and econometrics, the data are observed as time series that are potentially

* Corresponding author.
E-mail address: grice@uwaterloo.ca (G. Rice).

https://doi.org/10.1016/j.jmva.2018.07.001
0047-259X/

serially correlated or conditionally heteroscedastic. To give a few examples, PCA and the largest eigenvalues are often utilized in Markowitz portfolio optimization [29,30], and to model co-movements of markets and stocks as a barometer for risk [25,47], among many other applications. Moreover, with such time ordered data, it is often also of interest to determine before conducting such analyses whether or not the second order structure measured by the largest eigenvalues is homogeneous throughout the sample, or if it appears instead to exhibit one or more structural breaks. If the data under consideration consist of US macroeconomic indicators, for example, then the onset of a recession or the introduction of a new technology may be evidenced by instability in the largest eigenvalues of the covariance matrix. Additionally, PCA based forecasting methods might be improved if changes in the second order structure of the data are taken into account.

Motivated by these problems, in this paper we develop asymptotic results for the process of largest eigenvalues of the empirical covariance matrix based on an increasing proportion of the total sample. The asymptotic distribution of the eigenvalue process is established assuming the observations follow a high-dimensional factor model with common factors and errors satisfying a general weak dependence condition allowing for serial correlation and/or ARCH effects. A crucial distinction in the limiting behavior of the largest eigenvalue process is whether or not the common factor is the dominant component in the covariance matrix of the observations. If this is the case, then the largest eigenvalue diverges as $N \to \infty$, and we show that the process of largest eigenvalue estimates, suitably normalized, converges in distribution to a Gaussian process as $\min(N, T) \to \infty$. If the dependence on common factors is negligible, for instance if the cross sections are weakly correlated, which may be achieved by clustering or transforming the cross sections of a high-dimensional time series, then the largest eigenvalue of the covariance matrix is bounded, and the largest eigenvalue process has the usual Gaussian limit only when $N/\sqrt{T} \to 0$. We discuss the optimality of these rate conditions on N and T in detail in Theorem 3 and subsequent remarks, which basically state that without the given rate conditions one cannot obtain a Gaussian approximation for the centered largest eigenvalue process.

We further develop an application of these results to test for structural stability in linear factor models. These tests are based on maximally selected self normalized CUSUM statistics derived from the largest eigenvalue processes, which are shown to diverge in the presence of a common break in the mean or covariance matrix. An interesting technical challenge that must be faced when implementing these tests is the estimation of a normalizing sequence for the largest eigenvalue process which might diverge depending on the unknown rate of divergence of the largest eigenvalues. We show that this sequence can in general be consistently estimated using a simple kernel lag-window periodogram-type estimator, even when N is large in relation to T.

This work is inspired by, and builds upon, a number of recent contributions in change point analysis and structural break testing in linear factor models. With regards to testing for and estimating changes in the mean of high-dimensional linear factor models, we refer to Bai [5], who proposes a least squares change point estimator. Kim [26,27] extends this methodology to account for changes in linear trends in the presence of cross sectional dependence modeled by common factors. Horváth and Hušková [18] develop a test for detecting a change in the mean based on the CUSUM statistic. Li et al. [28] and Qian and Su [36] consider multiple structural breaks in panel data, and Kao et al. [22] consider break testing under cointegration.

Estimating and testing for changes in the covariance of scalar and vector-valued time series of a fixed dimension are considered in [2,13,42]. Kao et al. [23] develop a change point test based on the spectra and principal components of the covariance matrix in the finite-dimensional setting (N fixed). With regards to testing for changes in the factor structure in high-dimensional factor models, seminal work was conducted by Breitung and Eickmeier [9], who develop methodology for testing the constancy of factor loadings in individual cross sections using least squares regression of the cross section onto the principal component factors. Their test depends on estimating the number of common factors according to the information criterion developed in [6], and has been shown to be somewhat sensitive to the number of factors estimated. An alteration of their test to address this issue is developed in [44]. In [10,17], methods are proposed to test for changes in the factor loadings across a nonzero fraction of cross sections. In order that the estimation of the factors is asymptotically negligible, these tests generally assume that the factor loadings do not become degenerate as N increases, and that $\sqrt{T}/N \to 0$. In contrast, under their conditions our test only requires that $\min(N, T) \to \infty$.

The rest of the paper is organized as follows. In Section 2, we present our primary model and basic assumptions, as well as the main asymptotic results for the largest eigenvalue. Section 3 contains the details of applying the results of Section 2 to a change point problem, including asymptotic consistency results under the mean break and factor loading break alternatives. We derive consistent estimates of the aforementioned normalizing sequences as well as self-normalized statistics that bypass the need to perform such an estimation in Section 3. In Section 4, we discuss further practical details of the implementation of the proposed tests, and present the results of a Monte Carlo simulation study. Section 5 contains an application of the methodology developed in the paper to US treasury yield curve data. Analogous results for smaller eigenvalues are considered in Section 6. We provide some concluding remarks in Section 7, and all proofs of the technical results are collected in Section 8 as well as in an Online Supplement [19].

2. Models, assumptions, and main asymptotic results

We consider the linear factor model defined, for all $i \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$, by

$$X_{i,t} = \mu_i + \gamma_i \eta_t + e_{i,t}, \tag{1}$$

where $X_{i,t}$ denotes the ith coordinate or cross section vector-valued time series observed at time t, μ_i denotes the mean of the ith cross section, η_t denotes a real valued common factor with factor loadings γ_i, and $e_{i,t}$ denote the idiosyncratic errors. It is presumed that both the common factor and idiosyncratic errors may be serially correlated. As we develop asymptotics, we assume that the number of cross sections or dimension N and the sample size T tend jointly to infinity. We make the assumption that $\eta_t \in \mathbb{R}$ in order to simplify the presentation and the intuition behind the results presented below, and these results can be extended to the more general case of a vector-valued common factor and factor loading, which we discuss in Remark 2. We let $^\top$ denote the matrix transpose, and define the vectors $\mathbf{X}_t = (X_{1,t}, \ldots, X_{N,t})^\top \in \mathbb{R}^N$. We define, for all $u \in [1/T, 1]$,

$$\hat{\mathbf{C}}_{N,T}(u) = \frac{1}{\lfloor Tu \rfloor} \sum_{t=1}^{\lfloor Tu \rfloor} (\mathbf{X}_t - \bar{\mathbf{X}}_T)(\mathbf{X}_t - \bar{\mathbf{X}}_T)^\top,$$

to be the sample covariance matrix based on the proportion u of the sample, where $\bar{\mathbf{X}}_T = (\mathbf{X}_1 + \cdots + \mathbf{X}_T)/T$.

We first aim to study the processes derived from the K largest eigenvalues $\hat{\lambda}_1(u) \geq \cdots \geq \hat{\lambda}_K(u)$ of $\hat{\mathbf{C}}_{N,T}(u)$. At first we focus our attention on the process derived from the largest eigenvalue, and make the primary objective of this section to establish the weak convergence of $\hat{\lambda}_1(u)$ under model (1). Analogous results for processes derived from the smaller eigenvalues are provided in Section 6. We note that an alternative to using $\hat{\lambda}_i(u)$ is to use $\tilde{\lambda}_i(u) = (\lfloor Tu \rfloor/T)\hat{\lambda}_i(u)$, which are equivalent with the largest eigenvalues of the map defined, for all $u \in [0, 1]$, by

$$\tilde{\mathbf{C}}_{N,T}(u) = \frac{1}{T} \sum_{t=1}^{\lfloor Tu \rfloor} (\mathbf{X}_t - \bar{\mathbf{X}}_T)(\mathbf{X}_t - \bar{\mathbf{X}}_T)^\top.$$

Letting $\mathbf{C} = \mathrm{cov}(\mathbf{X}_t)$, we define the eigenvalues and eigenvectors of $\mathbf{C}$ by

$$\lambda_1 \mathfrak{e}_1 = \mathbf{C}\mathfrak{e}_1, \ldots, \lambda_N \mathfrak{e}_N = \mathbf{C}\mathfrak{e}_N, \tag{2}$$

where $\|\mathfrak{e}_1\| = \cdots = \|\mathfrak{e}_N\| = 1$, and $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^N$. Since N is allowed to increase with T, both the eigenvalues λ_i and eigenvectors $\mathfrak{e}_i$ may evolve as functions of N, i.e., $\lambda_i = \lambda_i(N)$. We suppress this dependence below for notational convenience. Throughout this paper, we make use of the following assumptions:

Assumption 1. There exists a positive integer K such that the eigenvalues $\lambda_1, \ldots, \lambda_K$ satisfy $\min_{i \in \{1,\ldots,K\}}(\lambda_i - \lambda_{i+1}) \geq c_0$ for some constant $c_0 > 0$ not depending on N.

Assumption 2. The common factor loadings satisfy that $|\gamma_i| \leq c_1$ for all $i \in \{1, \ldots, N\}$ with some $c_1 > 0$ not depending on N.

Assuming that the eigenvalues of $\mathbf{C}$ are distinct is necessary to derive a normal approximation for their estimates, and is a common assumption in the literature. We assume that the common factors and idiosyncratic errors satisfy a fairly general weak dependence condition.

Definition 1. We say that a stationary time series $\{\varepsilon_t : -\infty < t < \infty\}$ is an L^p-m-approximable Bernoulli shift with rate function χ if $\mathrm{E}\varepsilon_t = 0$, $\mathrm{E}\varepsilon_t^p < \infty$, and $\varepsilon_t = g(\nu_t, \nu_{t-1}, \ldots)$ for some measurable function $g : \mathbb{R}^\infty \to \mathbb{R}$ where $\{\nu_s : -\infty < s < \infty\}$ are independent and identically distributed random variables, and $\{\mathrm{E}(\varepsilon_t - \varepsilon_t^{(m)})^p\}^{1/p} = \chi(m)$ with $\varepsilon_t^{(m)} = g(\nu_t, \nu_{t-1}, \ldots, \nu_{t-m}, \nu^*_{t-m-1,t,m}, \nu^*_{t-m-2,t,m}, \ldots)$ and the $\nu^*_{i,j,\ell}$ are independent and identically distributed copies of ν_0.

The space of stationary processes that may be represented as Bernoulli shifts is quite large; we refer to [43] for a discussion. Examples include stationary ARMA, ARCH, and GARCH processes. The rate function describes the rate at which such processes can be approximated with sequences exhibiting a finite range of dependence. In many examples of interest, such as those listed above, the rate function may be taken to decay exponentially in the lag parameter.

Assumption 3.

(a) $\{\eta_t : -\infty < t < \infty\}$ is L^{12}-m-approximable with rate function $\chi_\eta(m) = c_2 m^{-\alpha_\eta}$ for constants $c_2 > 0$ and $\alpha_\eta > 1$, and $\mathrm{E}\eta_t^2 = 1$.
(b) The sequences $\{e_{i,t} : -\infty < t < \infty\}$, with $i \in \{1, \ldots, N\}$, are each L^{12}-m-approximable with rate functions $\chi_{e,i}(m) \leq c_3 m^{-\alpha_e}$ for constants $c_3 > 0$ and $\alpha_e > 1$. There exist constants c_4 and c_5 such that $0 < c_4 \leq \mathrm{E}e_{i,t}^2 = \sigma_{i,e}^2 \leq c_5 < \infty$.
(c) The sequences $\{\eta_t : -\infty < t < \infty\}$ and $\{e_{i,t} : -\infty < t < \infty\}$, for all $i \in \{1, \ldots, N\}$, are independent.

The least restrictive moment condition that could be assumed in order to obtain a normal approximation for the empirical eigenvalues is four moments. Our assumption of twelve moments comes from the fact that we apply a third order Taylor series expansion for the difference between the empirical eigenvalue process $\hat{\lambda}_i(u)$ and λ_i [15] and twelve moments are needed to get an upper bound for the highest order term that is uniform with respect to u. The condition in Assumption 3 that $\mathrm{E}\eta_t^2 = 1$ is nonrestrictive, as it makes the model (1) identifiable.

According to (1), we have that $\mathbf{C} = \boldsymbol{\gamma}\boldsymbol{\gamma}^\top + \boldsymbol{\Lambda}$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)^\top$, and $\boldsymbol{\Lambda}$ is an $N \times N$ diagonal matrix with $\sigma_{1,e}^2, \ldots, \sigma_{N,e}^2$ in the diagonal. As mentioned in the Introduction, the "strength" of the common factor is crucial in explaining the overall asymptotic behavior of λ_1 and $\hat{\lambda}_1(u)$. We consider two cases here separately: the case when $\|\boldsymbol{\gamma}\| \to \infty$ as $\min(N,T) \to \infty$, and the case when $\|\boldsymbol{\gamma}\| = O(1)$ as $\min(N,T) \to \infty$. In the first case, the common factor is the dominant term in the covariance matrix, and indeed $\hat{\lambda}_1(u)/\|\boldsymbol{\gamma}\|^2 \to 1$ in probability for all $u > 0$. In the latter case, according to Assumption 3(b), no one principal value of the covariance matrix dominates as $N \to \infty$, and the largest eigenvalue of $\mathbf{C}$ is bounded. This might be a reasonable assumption for high-dimensional time series from which the effect of the first few estimated common factors have been removed. In order to state the main result, we define $\xi_{i,t} = \mathfrak{e}_i^\top(\mathbf{X}_t - \mathrm{E}\mathbf{X}_0)(\mathbf{X}_t - \mathrm{E}\mathbf{X}_0)^\top \mathfrak{e}_i$.

Theorem 1. *Suppose (1) and Assumptions 1–3 hold, and $c \in (0,1]$. If $\|\boldsymbol{\gamma}\| \to \infty$, and*

$$\min\left\{\frac{N^2}{\|\boldsymbol{\gamma}\|^4 T^{1/2}}, \left(\frac{N^2}{T} + \frac{N}{T^{1/2}}\right)\frac{1}{\|\boldsymbol{\gamma}\|^2}\right\} \to 0, \tag{3}$$

then $T^{1/2}u\{\hat{\lambda}_1(u) - \lambda_1\}/(\|\boldsymbol{\gamma}\|^2\sigma_\eta) \xrightarrow{\mathcal{D}[c,1]} W(u)$, where $W(u)$ is a Wiener process, $\xrightarrow{\mathcal{D}[c,1]}$ denotes weak convergence in the Skorokhod topology on $[c,1]$, and

$$\sigma_\eta^2 = \sum_{t=-\infty}^{\infty} cov(\eta_0^2, \eta_t^2).$$

If instead $\|\boldsymbol{\gamma}\| = O(1)$ as $\min(N,T) \to \infty$, and

$$N/T^{1/2} \to 0, \tag{4}$$

then $T^{1/2}u\{\hat{\lambda}_1(u) - \lambda_1\}/\sigma_1 \xrightarrow{\mathcal{D}[c,1]} W(u)$, where

$$\sigma_1^2 = \sigma_1^2(N) = \sum_{t=-\infty}^{\infty} cov(\xi_{1,0}, \xi_{1,t}).$$

Remark 1. Theorem 1 shows that, under the given rate conditions relating N and T, the distribution of the largest eigenvalue process may be approximated by that of a Brownian motion. In the first case when $\|\boldsymbol{\gamma}\| \to \infty$, we note that if in addition $\|\boldsymbol{\gamma}\|^2$ is proportional to N, which would hold if $|\gamma_i| > 0$ for a nonzero proportion of the cross sections, or if γ_i were stochastic and independent of the common factor and idiosyncratic errors with a non-degenerate distribution, then (5) reduces to the condition that $\min(N,T) \to \infty$. Moreover, the norming sequence depends only on the long run variance of the common factors and the common factor loadings. In the case when $\|\boldsymbol{\gamma}\| = O(1)$, the norming sequence σ_1^2 is essentially the long run variance of the quadratic forms $\xi_{1,t}$, which is bounded under Assumptions 2 and 3.

In the above theorem we considered weak convergence of the largest eigenvalue process defined on an interval bounded away from the origin. One can obtain a similar result on the entire unit interval by assuming slightly stronger rate conditions on N and T, as demonstrated by the following theorem.

Theorem 2. *Suppose (1) and Assumptions 1–3 hold. If $\|\boldsymbol{\gamma}\| \to \infty$ as $\min(N,T) \to \infty$, and*

$$\min\left\{\frac{N^2(\ln T)^{1/3}}{\|\boldsymbol{\gamma}\|^4 T^{1/2}}, \left(\frac{N^2}{T} + \frac{N}{T^{1/2}}\right)\frac{(\ln T)^{1/3}}{\|\boldsymbol{\gamma}\|^2}\right\} \to 0, \tag{5}$$

then we have that $T^{1/2}u\{\hat{\lambda}_1(u) - \lambda_1\}/(\|\boldsymbol{\gamma}\|^2\sigma_\eta) \xrightarrow{\mathcal{D}[0,1]} W(u)$, If instead $\|\boldsymbol{\gamma}\| = O(1)$ as $\min(N,T) \to \infty$, and

$$N(\ln T)^{1/3}/T^{1/2} \to 0, \tag{6}$$

then $T^{1/2}u\{\hat{\lambda}_1(u) - \lambda_1\}/\sigma_1 \xrightarrow{\mathcal{D}[0,1]} W(u)$.

Remark 2. In the more general case where $\eta_t \in \mathbb{R}^p$ with factor loading matrix $\boldsymbol{\gamma} \in \mathbb{R}^{N\times p}$, we would require in Theorem 2 that the largest eigenvalue of $\boldsymbol{\gamma}^\top\boldsymbol{\gamma}$ is unique and tends to infinity in place of the condition that $\|\boldsymbol{\gamma}\| \to \infty$, or that the largest eigenvalue of $\boldsymbol{\gamma}^\top\boldsymbol{\gamma}$ is bounded in place of $\|\boldsymbol{\gamma}\| = O(1)$. The normalization in Theorem 1 by $\|\boldsymbol{\gamma}\|^2$ would be replaced by this eigenvalue.

When $\|\boldsymbol{\gamma}\|$ is bounded, conditions (4) and (6) require that the sample size T increases faster than the squared dimension N^2, which is quite restrictive. The case when N is proportional to T has received considerable attention in the probability and statistics literature. Assuming that $\hat{\mathbf{C}}_{N,T}(1)$ is based on independent and identically distributed entries, the distribution of $\hat{\lambda}_1(1)$ when properly standardized converges to a Tracy–Widom distribution [20]. For a survey of the theory of eigenvalues of large random matrices, we refer to [3]. The result in the following theorem shows that if (4) fails, then even in a somewhat simpler setting than considered above, $T^{1/2}\{\hat{\lambda}_1(1) - \lambda_1\}$ cannot be asymptotically Gaussian, and that, moreover, $\hat{\lambda}_1(1)$ asymptotically overestimates λ_1.

Theorem 3. *Suppose (1) and Assumptions 1–3 hold. In addition suppose that for each* $i \in \{1,\dots,N\}$, $\{e_{i,t} : t \in \mathbb{Z}\}$ *are independent and identically distributed, and* $\{\eta_t : t \in \mathbb{Z}\}$ *is a sequence of independent and identically distributed random variables, there exists a positive constant* c_6 *and* $\kappa > 2$ *such that* $\max_{i\in\{1,\dots,N\}} \mathrm{E}|e_{i,0}|^{3\kappa} \le c_6$ *and* $\mathrm{E}|\eta_0|^{3\kappa} < \infty$, *and* $N/T^{1/2} \to \infty$, $N^{1+2/\kappa}/T \to 0$. *Then, there exists a positive constant* c_7 *such that, as* $\min(N,T) \to \infty$, $\Pr[c_7 \le T^{1/2}\{\hat{\lambda}_1(1) - \lambda_1\}] \to 1$.

Theorem 3 basically shows then that in the absence of (4), one cannot obtain a Gaussian process for the centered eigenvalue process. Theorem 3 is proven in the Online Supplement [19].

3. Changepoint detection

Theorems 1 and 2 may be used to develop change point tests for the first and second order structure of high-dimensional data based on the largest empirical eigenvalue process. In order to formalize this, consider the model defined, for all $i \in \{1,\dots,N\}$ and $t \in \{1,\dots,T\}$, by

$$X_{i,t} = \{\mu_i + \delta_i \mathbf{1}(t \ge t^*)\} + \{\gamma_i + \psi_i \mathbf{1}(t \ge t^*)\}\eta_t + e_{i,t}, \tag{7}$$

in which the observations follow a factor model whose mean parameters and/or factor loadings might change at a common change point t^*. We would then like to test the null hypothesis $\mathcal{H}_0 : t^* > T$.

Under $\mathcal{H}_0$ the observed data follow (1). The largest eigenvalue of the covariance matrix is expected to be able to detect such changes for the following reasons. First of all, in the presence of a mean change the estimator $\bar{\mathbf{X}}_T$ converges to a linear combination of the means before and after the change point, which would asymptotically perturb the sample covariance matrices $\hat{\mathbf{C}}_{N,T}(u)$ and affect the largest eigenvalue process. Additionally, a large enough change in the loadings would also be expected to change the largest eigenvalue of the covariance matrix. Hence, a natural test statistic is to consider the size of the process defined, for all $u \in [c,1]$, by

$$\hat{B}_{T,1}(u) = T^{1/2}\, u\{\hat{\lambda}_1(u) - \hat{\lambda}_1(1)\}/\hat{\upsilon}_{1,T},$$

where $\hat{\upsilon}_{1,T}$ is a consistent estimator of either $\sigma_\eta^2\,\|\boldsymbol{\gamma}\|^2$ or σ_1^2. Such an estimator is studied in Section 3.1; see specifically Theorem 4.

Corollary 1. *Under the conditions of Theorems 1 and 4 below,* $\hat{B}_{T,1}(u) \xrightarrow{\mathcal{D}[c,1]} W^0(u)$, *where* W^0 *is a standard Brownian bridge.*

The continuous mapping theorem and Corollary 1 imply that

$$\sup_{u\in[c,1]} |\hat{B}_{T,1}(u)| \xrightarrow{\mathcal{D}} \sup_{u\in[c,1]} |W^0(u)|. \tag{8}$$

The limiting distribution on the right-hand side of (8) may be easily estimated for any value of the trimming parameter c using Monte Carlo simulation. An approximate test of size α of $\mathcal{H}_0$ is to reject if $\sup_{u\in[c,1]}|\hat{B}_{T,1}(u)|$ is larger than the α critical value of the estimated distribution. We study this test in finite samples in Section 4 below.

Remark 3. We would like to compare and contrast the conditions and strengths of this test to other tests available in the literature for detecting changes in the covariance structure of vector-valued observations. The tests developed in [9,10,17,44] are also cast in the setting of high-dimensional factor models, and rely on the condition that $\sqrt{T}/N \to 0$ in order that the estimation error incurred by estimating the common factors using PCA is asymptotically negligible. Their test is then based on detecting changes in the coefficients estimated from regressing the data onto these factor estimates, and importantly relies on a good estimate of the number of common factors. Our proposed test is instead based on measuring directly for changes in the covariance as measured in the largest eigenvalue without the need to estimate the number of common factors. In addition, their tests rely on the basic assumptions of [4], including their Assumption F3, which is effectively a stronger condition than $\|\boldsymbol{\gamma}\|^2 = O(N)$. Under this condition our test only requires that $\min(N,T) \to \infty$.

In the absence of $\sqrt{T}/N \to 0$, our proposed test is expected to work generally in two cases: one is in which $N/T^{1/2} \to 0$, and in this case our test does not require further conditions. This compares to the change point tests developed in [2,13,23] and [42] which each assume N is fixed. The aforementioned tests are quite sensitive to the dimension of the data since they rely on estimating and inverting the second order structure of the vectorized covariance matrix, which is akin to estimating and inverting a covariance matrix containing $N(N+1)/2$ elements. The second case is when the common factor is dominate, in which case again our method only requires that $\min(N,T) \to \infty$. An interesting open case is when the common factor is not dominate, and N and T are proportional. According to Theorem 3 a fundamentally different approach is required, perhaps relating to more modern Tracy–Widom limit laws.

3.1. Estimating the norming sequence

Consistent estimation of $\sigma_\eta^2\,\|\boldsymbol{\gamma}\|^2$ and σ_1^2 is required in order to apply Theorems 1 and 2 to test $\mathcal{H}_0$. Moreover, as it is unknown in practice whether $\|\boldsymbol{\gamma}\| \to \infty$ or $\|\boldsymbol{\gamma}\| = O(1)$, a desirable property is that the estimator be consistent to each possible limiting variance under the appropriate conditions. As these parameters represent the long run variance of the quadratic forms $\xi_{i,t}$, we propose a standard nonparametric estimator of the spectral density at frequency zero. We define $\hat{\mathfrak{e}}_i$ for each $i \in \{1,\dots,N\}$ by $\hat{\lambda}_i(1)\hat{\mathfrak{e}}_i = \hat{\mathbf{C}}_{N,T}(1)\hat{\mathfrak{e}}_i$, and let $\hat{\xi}_{i,t} = \{\hat{\mathfrak{e}}_i^\top(\mathbf{X}_t - \bar{\mathbf{X}}_T)\}^2$.

Let J be a kernel/weight function that is positive, continuous and symmetric about the origin in $\mathbb{R}$ with bounded support that satisfies $J(0) = 1$. Examples of such functions include the Bartlett and Parzen kernels; further examples and discussion may be found in [39]. We define the estimator $\hat{v}_{1,T}^2$ for either $\sigma_\eta^2 \|\boldsymbol{\gamma}\|^2$ or σ_1^2 by

$$\hat{v}_{1,T}^2 = \sum_{s=-N+1}^{N-1} J(s/h)\, \hat{r}_{1,s}, \tag{9}$$

where h denotes a smoothing bandwidth parameter, and

$$\hat{r}_{1,s} = \begin{cases} \dfrac{1}{T-s} \displaystyle\sum_{t=1}^{T-s} (\hat{\xi}_{1,t} - \bar{\xi}_{1,T})(\hat{\xi}_{1,t+s} - \bar{\xi}_{1,T}) & \text{if } s \geq 0, \\ \dfrac{1}{T-|s|} \displaystyle\sum_{t=-s}^{T} (\hat{\xi}_{1,t} - \bar{\xi}_{1,T})(\hat{\xi}_{1,t+s} - \bar{\xi}_{1,T}) & \text{if } s < 0, \end{cases}$$

with $\bar{\xi}_{1,T} = (\hat{\xi}_{1,1} + \cdots + \hat{\xi}_{1,T})/T$. The following theorem establishes the consistency of this estimate, and is proven in the Online Supplement [19].

Theorem 4. *Suppose $\mathcal{H}_0$ and Assumptions 1–3 hold. If $\|\boldsymbol{\gamma}\| \to \infty$ as $\min(N,T) \to \infty$, and*

$$h = h(T) \to \infty \quad \text{with} \quad h\left(\frac{N}{\|\boldsymbol{\gamma}\|^2 T^{1/2}} + \frac{N^{3/2}}{\|\boldsymbol{\gamma}\|^4 T^{1/2}}\right) \to 0, \tag{10}$$

then, as $\min(N,T) \to \infty$, $\hat{v}_{1,T}^2/(\sigma_\eta^2 \|\boldsymbol{\gamma}\|^4) \xrightarrow{P} 1$. If instead $\|\boldsymbol{\gamma}\| = O(1)$, then if

$$hN^{3/2}/T^{1/2} \to 0, \tag{11}$$

$\hat{v}_{1,T}^2/\sigma_1^2 \xrightarrow{P} 1$ *as* $\min(N,T) \to \infty$.

Remark 4. Theorem 4 provides rate conditions on the bandwidth parameter h that involve both N and T, as well as $\|\boldsymbol{\gamma}\|$ in the case when it diverges. If $\|\boldsymbol{\gamma}\|^2$ is proportional to N, then (10) reduces to the condition that $h/T^{1/2} \to 0$, which is a standard condition in nonparametric spectral density and long run variance estimation. On the other hand, condition (11) is somewhat less appealing since it differs with the apparent optimal condition based on Theorem 2 that $hN/T^{1/2} \to 0$ by a factor of $N^{1/2}$. This may be avoidable by an alternative method of proof, but the given condition still allows for a similar divergence rate of N with respect to T. Theorem 4 is proven in the Online Supplement [19].

3.2. Self-normalized/ratio statistics

One way of avoiding the estimation of the long run variance parameter is to consider self-normalized statistics. The simple idea is that a ratio of two functionals of $\hat{\lambda}_1(u)$ can be constructed in such a way that both its limiting distribution does not depend on the scaling constants σ_η nor σ_1, and the corresponding test statistic still has power. Self-normalization has been utilized fairly extensively in the context of change point analysis, see, e.g., [16] and [46], and these results among others on the topic are summarized in the review [38].

Towards defining a self-normalized statistic based on the largest eigenvalue process, we define $\tilde{\lambda}_1^{(1)}(v,s)$ to be the largest eigenvalue of the matrix defined for all $v, s \in \{1,\dots,T\}$ by

$$\hat{\mathbf{C}}_{N,T}^{(1)}(v,s) = \frac{1}{T}\sum_{t=1}^{v} (\mathbf{X}_t - \bar{\mathbf{X}}_s)(\mathbf{X}_t - \bar{\mathbf{X}}_s)^\top,$$

where $v \in \{1,\dots,s\}$, and $\bar{\mathbf{X}}_s = (\mathbf{X}_1 + \cdots + \mathbf{X}_s)/s$. Similarly we define $\tilde{\lambda}_1^{(2)}(v,s)$ to be, for $v, s \in \{1,\dots,T\}$, the largest eigenvalue of

$$\hat{\mathbf{C}}_{N,T}^{(2)}(v,s) = \frac{1}{T}\sum_{t=v+1}^{T} (\mathbf{X}_t - \tilde{\mathbf{X}}_s)(\mathbf{X}_t - \tilde{\mathbf{X}}_s)^\top,$$

where $v \in \{s,\dots,T\}$, and $\tilde{\mathbf{X}}_s = (\mathbf{X}_{s+1} + \cdots + \mathbf{X}_T)/(T-s)$. We then base the test statistic on

$$Q_{N,T}^{(1)}(s) = \max_{1 \leq v \leq s} |\tilde{\lambda}_1^{(1)}(v,s) - v\tilde{\lambda}_1^{(1)}(s,s)/s| \quad \text{and} \quad Q_{N,T}^{(2)}(s) = \max_{s \leq v < T} |\tilde{\lambda}_1^{(2)}(v,s) - (T-v)\tilde{\lambda}_1^{(2)}(s,s)/(T-s)|.$$

Roughly speaking, given a point s that partitions the total sample into two subsamples, we calculate functionals of the largest eigenvalue processes in each subsample in $Q_{N,T}^{(1)}$ and $Q_{N,T}^{(2)}$. Under $\mathcal{H}_0$, the limit distribution of the ratios of these quantities can be obtained. Letting $c \in (0, 1/2)$, we define

$$Q_{N,T} = Q_{N,T}^{(c)} = \max\left\{ \frac{\max_{s\in\{1,\ldots,\lfloor T/2\rfloor\}} Q_{N,T}^{(2)}(s)}{\max_{s\in\{\lfloor cT\rfloor,\ldots,\lfloor T/2\rfloor\}} Q_{N,T}^{(1)}(s)}, \frac{\max_{s\in\{\lfloor T/2\rfloor,\ldots,T\}} Q_{N,T}^{(1)}(s)}{\max_{s\in\{\lfloor T/2\rfloor,\ldots,\lfloor(1-c)T\rfloor\}} Q_{N,T}^{(2)}(s)} \right\},$$

and

$$Q = Q^{(c)} = \max\left[\frac{\sup_{0\le x\le 1/2}\sup_{x\le u\le 1}|W(1)-W(u)-(1-u)\{W(1)-W(x)\}/(1-x)|}{\sup_{c\le x\le 1/2}\sup_{0\le u\le x}|W(u)-uW(1)/x|}, \frac{\sup_{1/2\le x\le 1}\sup_{0\le u\le x}|W(u)-uW(1)/x|}{\sup_{1/2\le x\le 1-c}\sup_{x\le u\le 1}|W(1)-W(u)-(1-u)\{W(1)-W(x)\}/(1-x)|}\right].$$

Theorem 5. *If the conditions of Theorem 1 are satisfied, and either $\|\boldsymbol{\gamma}\| \to \infty$ and (3) holds, or $\|\boldsymbol{\gamma}\| = O(1)$ and (4) holds, then $Q_{N,T} \xrightarrow{D} Q$, as $\min(N,T) \to \infty$.*

Theorem 5 is proven in the Online Supplement [19]. The limiting distribution of Q is free of nuisance parameters, and its quantiles can be produced for a given value of c via Monte Carlo simulation. There are several other ways to define such self normalized statistics, for example

$$Q_{N,T}^{*} = \max\left\{ \max_{s\in\{\lfloor cT\rfloor,\ldots,\lfloor T/2\rfloor\}} Q_{N,T}^{(2)}(s)/Q_{N,T}^{(1)}(s), \max_{s\in\{\lfloor T/2\rfloor,\ldots,\lfloor(1-c)T\rfloor\}} Q_{N,T}^{(1)}(s)/Q_{N,T}^{(2)}(s)\right\}$$

could also be used. Moreover, instead of calculating maximum differences in $Q_{N,T}^{(i)}(s)$ $i = 1{,}2$, one might also consider the sum of squared differences.

3.3. Consistency under alternatives

We now turn our attention to studying the consistency of tests for $\mathcal{H}_0$ based on $\sup_{u\in[c,1]}|\hat{B}_{T,1}(u)|$ under the mean break and factor loading break alternatives. We assume that the change does not occur too close to the end points of the sample, i.e. for some $\theta \in (0, 1)$

$$t^{*} = \lfloor T\theta\rfloor \tag{12}$$

First we consider the case of a break in the mean, i.e., the model for $i \in \{1,\ldots,N\}$ and $t \in \{1,\ldots,T\}$

$$X_{i,t} = \{\mu_i + \delta_i \mathbf{1}(t \ge t^{*})\} + \gamma_i\eta_t + e_{i,t}, \tag{13}$$

holds. Let $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_N)^{\top}$ and assume

$$\lim_{\min(N,T)\to\infty} T^{1/2}\|\boldsymbol{\delta}\|/\|\boldsymbol{\gamma}\|^2 = \infty. \tag{14}$$

Theorem 6. *Under (13), Assumptions 1–3, and assuming that (3), (12), and (14) are satisfied, then we have that $\sup_{u\in[c,1]}|\hat{B}_{T,1}(u)| \xrightarrow{P} \infty$.*

We note that assumptions (12) and (14) also appeared in [18] where their optimality is discussed. It is clear if N is large, relatively small changes can be detected by $\hat{\lambda}_1(u)$. Moreover, we note that condition (14) does not require that there are changes in all cross sections. As a consequence of the proof of Theorem 6, it follows that

$$\max_{i\in\{2,\ldots,K\}}\sup_{u\in[0,1]} T^{1/2}|\hat{\lambda}_i(u) - \hat{\lambda}_i(1)| = O_P(1),$$

i.e., a change in the mean is asymptotically entirely captured by the largest eigenvalue of the partial covariance matrices.

The condition (14) suggests how a local change in the mean alternative may be considered. For example, if $\delta_1 = \cdots = \delta_N = \delta(N,T)$ and $\gamma_1 = \cdots = \gamma_N = \gamma$, γ being fixed, we need that $(T/N)^{1/2}|\delta(N,T)| \to \infty$ for (14) to hold, which describes at what rate $\delta(N,T)$ may tend to zero while maintaining consistency.

Next we consider the model for $i \in \{1,\ldots,N\}$ and $t \in \{1,\ldots,T\}$ defined by

$$X_{i,t} = \mu_i + \{\gamma_i + \psi_i \mathbf{1}(t \ge t^{*})\}\eta_t + e_{i,t}, \tag{15}$$

i.e., the means of the panels remain the same but the loadings change at time t^{*}. Let $\boldsymbol{\psi} = (\psi_1,\ldots,\psi_N)^{\top}$.

Theorem 7. *Under (15), Assumptions 1–3, and assuming that (3), (12) and*

$$\lim_{\min(N,T)\to\infty} \frac{(1-\theta)[\|\boldsymbol{\psi}\|^2 + 2|\boldsymbol{\psi}^\top\boldsymbol{\gamma}|] + (\boldsymbol{\psi}^\top\boldsymbol{\gamma})^2/\|\boldsymbol{\psi}\|^2}{\|\boldsymbol{\gamma}\|^2 + \max_{1\le i\le N}\sigma_i^2} > 1 \tag{16}$$

hold, then $\sup_{u\in[c,1]}|\hat{B}_{T,1}(u)| \stackrel{P}{\to} \infty$*, as* $\min(N,T)\to\infty$.

Roughly speaking, it is possible that the covariance might change on a subspace that is orthogonal to the first eigenvector (or more generally the first K eigenvectors), and then if this change is not sufficiently large, the first eigenvalue cannot have power to detect it. Condition (16) is sufficient to imply that this does not occur. One can show in addition that $Q_{N,T}\to\infty$ in probability under the conditions of Theorems 6 and 7.

3.4. Differentiating between level shifts and changes in the common factor loadings

Under model (7), a rejection of $\mathcal{H}_0$ based on the supremum of the largest eigenvalue process may be caused by either a change in the cross sectional means, or a change in the common factor loadings of the cross sections, and in some cases it may be of interest to differentiate between these potential causes. We leave the development of a complete theory to address this problem for future research, but a simple idea to accomplish this is to, in the case of a rejection of $\mathcal{H}_0$, perform the test additional times after adjusting for potential changes in the mean and/or common factor loadings. Let

$$\bar{\mathbf{X}}^*_{T,t} = \begin{cases} \dfrac{1}{\hat{t}^*}\displaystyle\sum_{t=1}^{\hat{t}^*}\mathbf{X}_t & \text{if } t\in\{1,\ldots,\hat{t}^*\},\\[2ex] \dfrac{1}{T-\hat{t}^*}\displaystyle\sum_{t=\hat{t}^*+1}^{T}\mathbf{X}_t & \text{if } t\in\{\hat{t}^*+1,\ldots,T\},\end{cases}$$

where $\hat{t}^*$ is the least squares change point estimator for a change in the mean defined in Section 3 of [5], and define $\mathbf{X}^*_t = \mathbf{X}_t - \bar{\mathbf{X}}^*_{T,t}$. If the test no longer rejects when applied to the observations $\mathbf{X}^*_t$, then presumably the cause of the original rejection was a change in the cross sectional means. Otherwise, we suspect there is a change to the factor loadings. Of course in this latter case it is possible that there is a change in both the means and common factor loadings. In order differentiate these possibilities, let $\hat{\lambda}_{c,1}$, $\hat{\mathfrak{e}}_{c,1}$ denote the largest eigenvalue and eigenvector, respectively, of the sample covariance matrix of $\mathbf{X}^*_t$, with $t\in\{1,\ldots,\hat{t}^*\}$, and similarly define $\hat{\lambda}_{c,2}$, $\hat{\mathfrak{e}}_{c,2}$ based on the sample covariance matrix of $\mathbf{X}^*_t$ with $t\in\{\hat{t}^*+1\ldots,T\}$. We then define $\mathbf{X}^{**}_t = \mathbf{X}_t - \langle\mathbf{X}_t,\hat{\mathfrak{e}}_{c,1}\rangle\hat{\mathfrak{e}}_{c,1}$ for all $t\in\{1,\ldots,\hat{t}^*\}$, and $\mathbf{X}^{**}_t = \mathbf{X}_t - \langle\mathbf{X}_t,\hat{\mathfrak{e}}_{c,2}\rangle\hat{\mathfrak{e}}_{c,2}$ for $t\in\{\hat{t}^*+1,\ldots,T\}$. If the test rejects when applied to $\mathbf{X}^{**}_t$, then we conclude there is a change in the means and common factor loadings.

4. Finite-sample performance

In order to demonstrate how the above results are manifested in finite samples, we present here the results of a Monte Carlo simulation study involving several different data generating processes (DGPs) that follow (7). All simulations were carried out in the R programming language (R Development Core Team [37]). In order to compute the long run variance estimate $\hat{v}^2_{1,T}$ defined in (9), we used the "sandwich" package (see Zeileis [45]), in particular the "kernHAC" function. The Parzen kernel with corresponding empirical bandwidth defined in [1] computed from the random variables $\hat{\xi}_{i,t}$ were employed. An important consideration in structural break testing brought to light in [34] and [40] is that of non-monotonic power of the test relative to increasing sizes of the change. In order to account for this, we replaced the mean estimates $\bar{\mathbf{X}}$ used in the definition of the long run variance estimator with $\bar{\mathbf{X}}^*_{T,t}$ defined in Section 3.4.

4.1. Empirical size

We begin by presenting the results on the empirical size of the tests for stability based on the largest eigenvalue by considering two examples of synthetic data generated according to model (1). We use the notation $Y_i\sim Y$ to denote that the sequence of random variables Y_i are independent and identically distributed with distribution Y. Let $\mathcal{N}_{i,t}(0,1)$, $i\in\{0,1,\ldots\}$, and $t\in\mathbb{Z}$ denote iid standard normal random variables, and let $AR_i(1,p)$ $i\in\{0,1,\ldots\}$ denote independent autoregressive one processes with parameter p based on standard normal errors. We generated observations $X_{i,t}$ according to (1) and the DGPs

(IID): $\eta_t = \mathcal{N}_{0,t}(0,1)$, $e_{i,t} = s_i\mathcal{N}_{i,t}(0,1)$, $s_i\sim\mathcal{U}(.8,1.2)$, $\gamma_i\sim\mathcal{N}(0,1)$;
(AR1): $\eta_t = AR_0(1,.5)$, $e_{i,t} = s_iAR_i(1,.5)$, $s_i\sim\mathcal{U}(.8,1.2)$, $\gamma_i\sim\mathcal{N}(0,1)$.

The purpose of choosing random parameters s_i, which define the standard deviations of the idiosyncratic errors, and γ_i is two fold. Firstly, this forces Assumption 1 to hold. Secondly, this choice highlights that the methodology is relatively robust to variations in the parameter values.

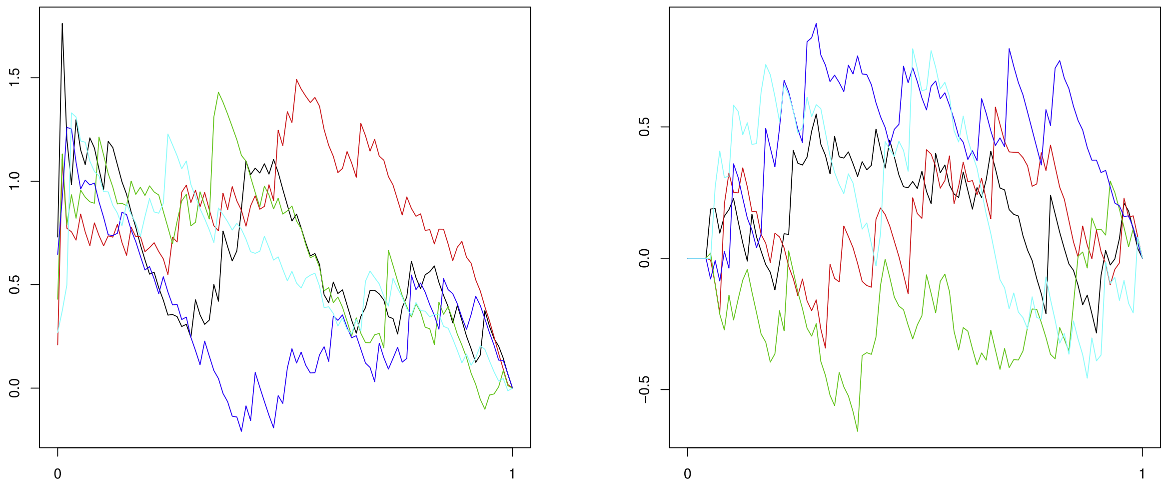

Fig. 1. The left panel illustrates five simulated paths of $\hat{B}_{T,1}(u)$ when $N = 20$ and $T = 100$ under (IID), and the right panel illustrates five simulated paths of $\tilde{B}_{T,1}(u)$ under the same conditions with $c = 0.05$.

Five simulated paths of the process $\hat{B}_{T,1}(u)$ are shown in the left hand panel of 1 when $T = 100$ and $N = 20$, under IID. The most notable feature is that each process always starts with a spike near the origin, i.e., $\hat{\lambda}_i(u)$ is much larger than $\hat{\lambda}_i(1)$ when u is small. The reason for this is that, when u is small, $\hat{\lambda}_i(u)$ is computed from a matrix that is low rank, and hence will tend to be closer to the norm of the observation vectors, which is on the order of N, than the eigenvalue that it being estimated. This problem is ameliorated when N decreases or T increases, but significantly affects the results for many practical values of N and T.

In order to correct for this, we define, for all $u \in (c, 1]$,

$$\tilde{B}_{T,1}(u) = \hat{B}_{T,1}(u) - (1-u)\hat{B}_{T,1}(c)/(1-c)$$

for a trimming parameter $0 < c < 1$. It follows then under the conditions of Theorem 1 that

$$\sup_{u\in[c,1]} |\tilde{B}_{T,1}(u)| \xrightarrow{D} \sup_{u\in[c,1]} |W^0(u) - (1-u)W^0(c)/(1-c)|, \tag{17}$$

and the limit on the right-hand side can be approximated by Monte Carlo simulation. Five corresponding paths of $\tilde{B}_{T,1}(u)$ are illustrated in the right panel of Fig. 1, with $c = 0.05$.

Table 1 contains the percentages of the test statistics $\sup_{u\in[c,1]}|\tilde{B}_{T,1}(u)|$ and $Q_{N,T}^{(c)}$ that are larger than the 10%, 5%, and 1% critical values of the distribution on the right-hand side of (17) and $Q^{(c)}$. The results can be summarized as follows:

1. When T is small ($T = 50$), then the size of the test may be inflated by two sources. One of them is the spiked effect, and this is particularly pronounced when c is small and N is large. If the temporal dependence in the data is low, then increasing c can allow the test to achieve good size even for small T and relatively large N. However, strong temporal dependence can cause size inflation for small T that cannot be accounted for by increasing c. In general as long as the sample size is at least larger than or equal to the dimension we recommend $c = 0.1$, which we use in the application below.
2. Another source of size inflation that is present for larger values of T may be attributed to estimating the variance under the alternative of a break in the mean. This may be improved by considering alternative variance estimation approaches, such as those developed in [24].
3. The difference in the results between the IID and AR1 DGP's were small for larger values of T ($T = 100, 200$), indicating that the nonparametric long run variance estimation performs relatively well with large enough samples.
4. For $T = 200$, the empirical sizes are close to nominal in all cases.
5. The self normalized statistic $Q_{N,T}^{(c)}$ exhibited similar behavior regarding empirical size, and even some improvement when N is large relative to T.

4.2. Empirical power

In order to study the power of our test under both the mean break and loading break alternatives, we considered two processes that satisfy (7) with $t^* = T\theta$ with $\theta \in (0, 1)$. Throughout the simulations below, we set $t^* = T/2$, i.e., the break was

Table 1
Empirical sizes with nominal levels of 10%, 5%, and 1% in both the independent (IID) and dependent (AR1) cases based on the statistics $\sup_{u\in[c,1]}|\tilde{B}_{T,1}(u)|$, and $Q_{N,T}^{(c)}$.

DGP:		IID						AR1					
		$c = 0.05$			$c = 0.1$			$c = 0.05$			$c = 0.1$		
N	T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
$\sup_{u\in[c,1]}\|\tilde{B}_{T,1}(u)\|$													
10	50	18.1	11.2	3.8	8.8	4.9	1.8	26.7	18.4	10.0	24.7	17.9	8.4
	100	8.3	3.5	0.7	9.2	3.6	0.7	17.1	10.3	3.4	9.2	3.6	0.7
	200	8.7	4.1	0.7	8.6	4.3	1.0	11.7	5.7	2.0	10.4	5.1	1.6
20	50	18.6	12.3	5.5	9.5	4.8	0.7	23.7	16.5	8.0	25.8	17.8	8.7
	100	8.5	3.6	0.6	9.1	4.5	0.3	14.9	9.4	3.4	14.9	9.0	3.7
	200	8.4	4.2	0.6	8.8	3.3	0.5	11.8	6.5	2.0	12.4	6.8	1.5
50	50	23.3	13.7	5.3	10.2	3.9	0.7	24.8	17.3	8.8	24.4	18.6	4.9
	100	8.8	3.5	0.6	9.0	4.2	1.0	17.8	11.6	4.0	15.3	8.5	3.5
	200	10.0	5.0	1.3	8.9	3.8	0.5	13.0	7.1	2.1	12.2	6.4	1.7
$Q_{N,T}^{(c)}$													
10	50	15.7	10.8	4.8	12.5	8.5	4.0	21.2	15.8	8.8	17.7	13.4	7.1
	100	10.7	5.9	2.0	9.8	6.7	1.7	14.6	9.2	4.5	12.7	8.2	3.4
	200	6.3	3.6	0.9	6.9	3.6	0.7	10.3	5.7	2.3	10.0	6.2	2.2
20	50	14.7	9.2	4.0	12.4	8.4	3.2	20.4	14.7	7.5	17.4	13.3	7.6
	100	10.3	6.7	2.1	10.0	6.7	2.3	11.8	7.9	3.5	14.3	10.3	3.7
	200	7.2	3.4	0.8	7.6	4.0	0.6	11.8	7.1	3.0	9.2	6.2	1.5
50	50	13.9	9.7	4.1	12.7	7.2	3.0	21.8	16.3	10.4	17.6	12.3	5.5
	100	9.6	5.2	1.2	9.4	5.8	1.6	14.9	10.0	4.8	14.5	9.9	4.8
	200	7.7	4.1	1.1	6.5	3.7	1.3	9.9	6.0	2.4	9.2	6.0	1.6

in the middle of the sample. We also studied the situation in which breaks occurred towards the endpoints of the sample. The results in those cases tended to be worse, but not more so than is typical in these problems. We define the DGPs for $i \in \{1,\ldots,N\}$ and $t \in \{1,\ldots,T\}$ by

MB(δ): $X_{i,t} = \delta_i \mathbf{1}(t \geq T/2) + \gamma_i\eta_t + e_{i,t}$, where $\delta_i \sim \mathcal{U}(-\delta,\delta)$;
LB(Δ): $X_{i,t} = \{\gamma_i + \psi_i \mathbf{1}(t \geq T/2)\}\eta_t + e_{i,t}$, where $\psi_i \sim \mathcal{N}(0,\Delta^2)$.

In each case we take the other terms in (7), i.e., the idiosyncratic errors, common factor, and factor loadings, to satisfy AR1. The results improve when the data are IID. We let the parameters δ and Δ vary between 0 and 3 at increments of 0.5, and let $N \in \{10, 20, 50\}$, and $T \in \{50, 100, 200\}$. The results are displayed in terms of power curves in Fig. 2 for the statistic $\sup_{u\in[0.1,1]}|\tilde{B}_{T,1}(u)|$, and Fig. 3 for $Q_{N,T}^{(0.1)}$ when the significance level of the test was fixed at 5%. We summarize the results as follows.

Mean break

1. For each value of T and N that we considered there is a substantial gain in power for δ exceeding 1.5 with the statistic $\sup_{u\in[0.1,1]}|\tilde{B}_{T,1}(u)|$. We note that data generated according to AR1 have cross-sectional standard deviations of on average 1.6, and, when $\delta = 2$, the average squared size of the change in the mean of each cross section is 1.33. Thus testing based on the largest eigenvalue seemed fairly sensitive to detect changes in the mean. The statistic $Q_{N,T}$ by contrast was less sensitive for detecting changes in the mean.
2. Due to the estimation of the variance under a mean break, the test based on $\sup_{u\in[c,1]}|\tilde{B}_{T,1}(u)|$ exhibited monotonic power.
3. Increasing T with fixed N improved the empirical power, as expected, and the same was observed when T was fixed and N increased. The latter occurrence is likely attributable to the fact that as N increases, changes in the mean occur in more cross sections, and the size is inflated in these cases due to the spiked effect.

Loading break

1. In the case of a break in the factor loadings, even smaller changes relative to the size of the standard deviation ($\Delta = 1$) of the idiosyncratic errors resulted in dramatic increases in power.
2. We noticed that for smaller values of T ($T = 50$) the power seemed to level off for larger breaks in the common factors, and never reached more than 90%.
3. For larger T ($T \in \{100, 200\}$), the power approached 1 at a much faster rate for breaks in the factor loadings, and this occurrence seemed to be independent of the value of N.

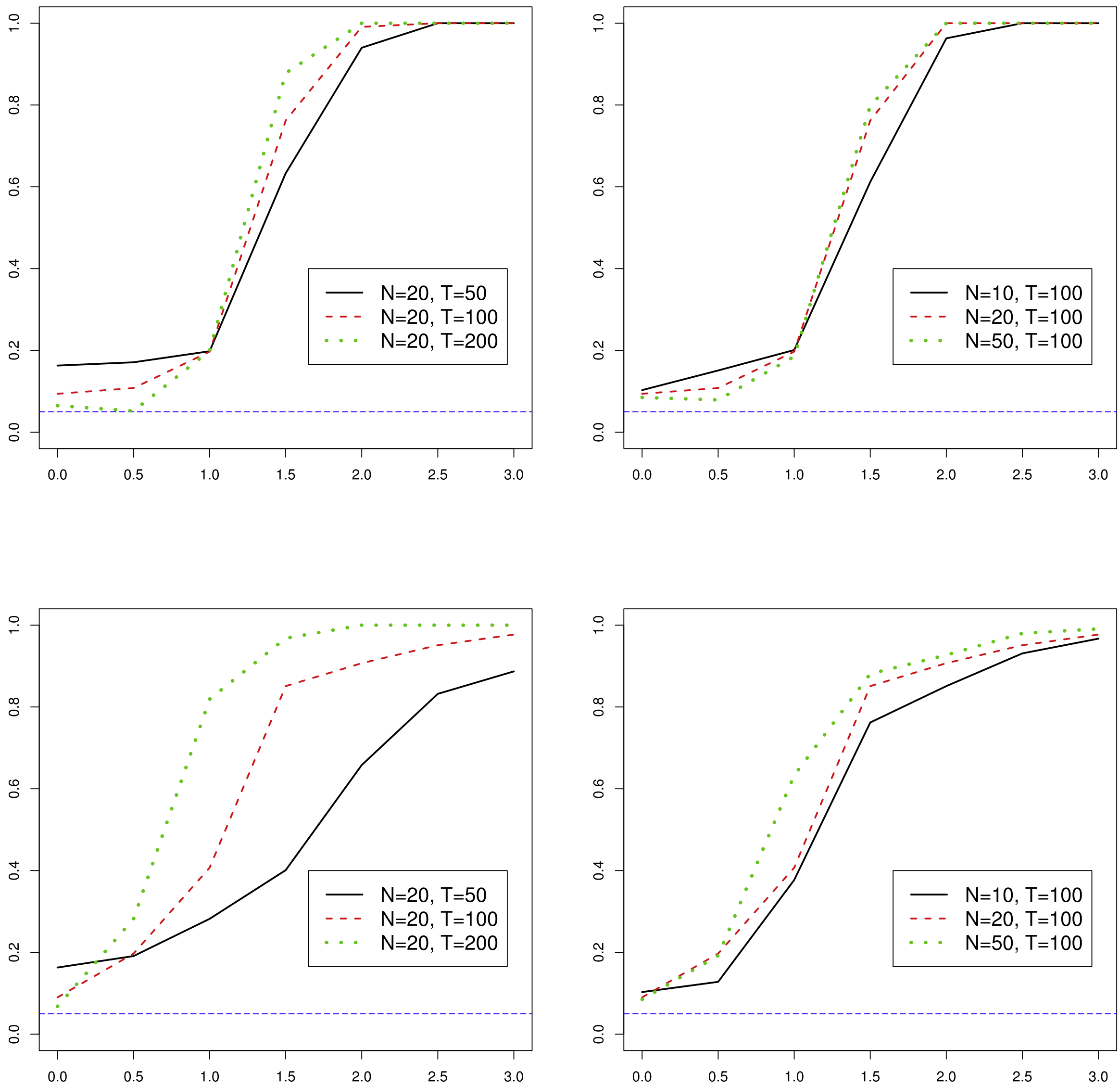

Fig. 2. The top two panels contain power curves generated from data following MB(δ) based on $\sup_{u\in[0.1,1]}|\tilde{B}_{T,1}(u)|$ for varying N and T as a function of δ. The bottom two panels contain similar curves generated from data following LB(Δ). The horizontal axis measures δ, Δ, and the vertical axis measures the empirical power when the significance level is fixed at 5%, indicated by the bottom horizontal (blue) line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Increasing N resulted in reduced power in this case, although the effects of changing N were not particularly pronounced.
5. The statistic $Q_{N,T}$ outperformed $\sup_{u\in[c,1]}|\tilde{B}_{T,1}(u)|$ with regards to detecting changes in the loadings. The two statistics thus exhibited complementary strengths.

5. Application to US yield curve data

Following [44], we consider an application of our methodology to test for structural breaks in US Treasury yield curve data considered in Gürkaynak et al. [14], which are available at http://www.federalreserve.gov/econresdata, and which the authors graciously maintain. The data consist of estimates of the yields for fixed interest securities with maturities between one and thirty years with one year increments ($N = 30$). We studied a portion of this data set spanning from November 25, 1985 to June 5, 2017 that we further reduced from daily to approximately monthly observations by considering only the data from each 22nd day, resulting in time series of length 356 for each maturity. Fig. 4 illustrates the yield curves corresponding to 1, 5, 10, and 30 year maturities for a subset of the data.

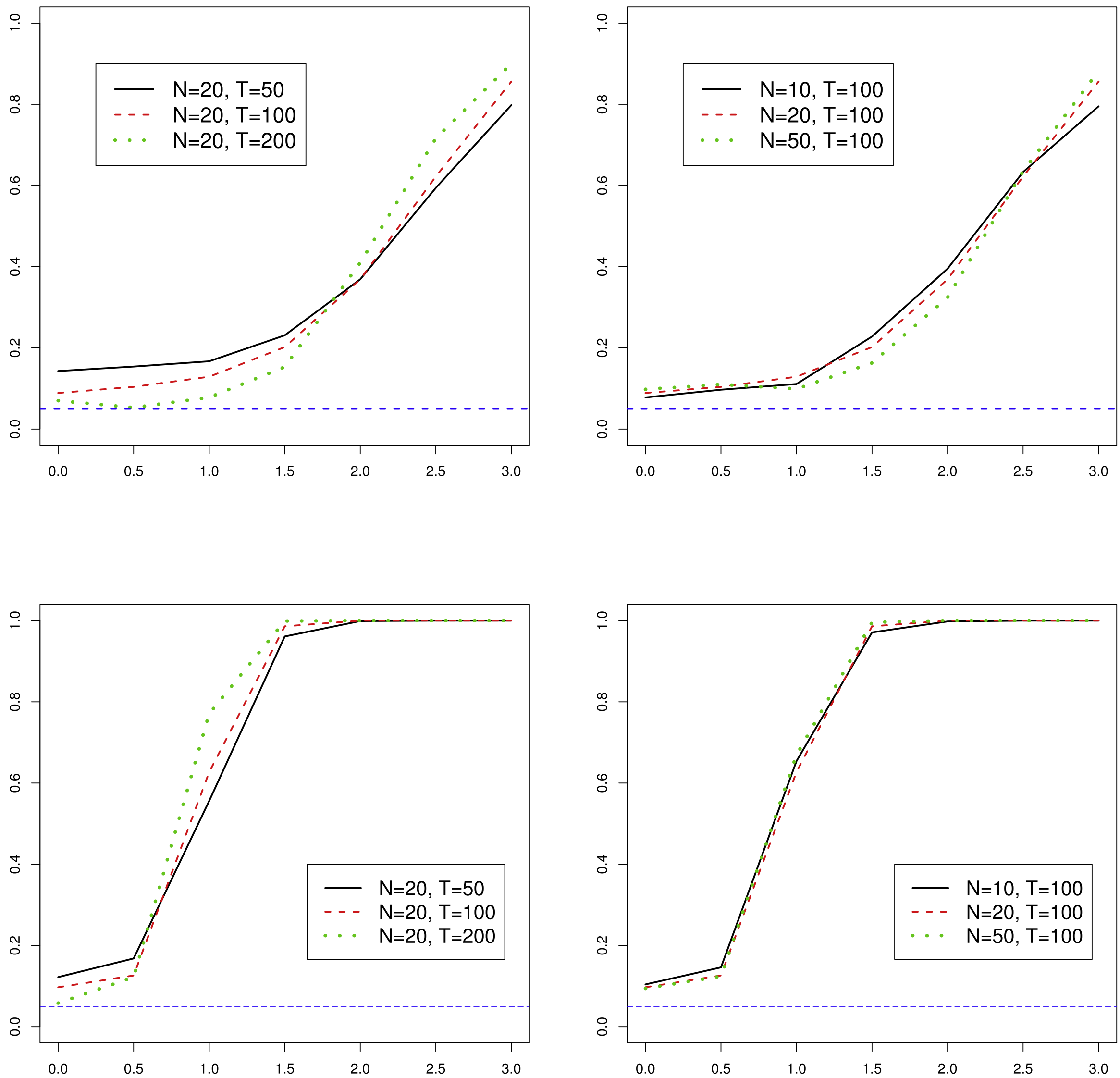

Fig. 3. The top two panels contain power curves generated from data following MB(δ) based on $Q_{N,T}^{(0.1)}$ for varying N and T as a function of δ. The bottom two panels contain similar curves generated from data following LB(Δ). The horizontal axis measures δ, Δ, and the vertical axis measures the empirical power when the significance level is fixed at 5%, indicated by the bottom horizontal (blue) line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In order to remove the effects of stochastic trends, and to allow for a comparison of our results to [44], we first differenced each series. The eigenvalues calculated from the covariance matrix calculated from the full, first differenced, sample $\hat{\mathbf{C}}_{30,356}(1)$ are displayed in the right-hand panel of Fig. 4, from which it is clear that the largest eigenvalue is substantially greater than the others. This suggests the validity of the single dominant common factor model that is the subject of the first half of Theorem 1.

We applied the hypothesis test for stability of the largest eigenvalue based on $\sup_{u\in[c,1]}|\tilde{B}_{T,1}(u)|$ with trimming parameter $c = 0.1$ to sequential blocks of the first differenced data of length 5 years, corresponding to 60 monthly observations in each sample ($T = 60$). The first block contained data spanning from November, 1985 to October, 1999, and the last block contained data spanning from July, 2012 to June, 2017, which constituted a total of 296 tests. The p-value from each test is plotted against the end date of the corresponding 5 year block in Fig. 5.

One notices when looking at Fig. 5 several points at which the largest eigenvalue of the sample covariance matrix appeared to be unstable. The first appears in the years from 1986–1991. More pronounced instability appears in the years following 2001, presumably spurred by the 9/11 attacks and the bursting of the dot-com bubble, and again in the years leading up to the subprime crisis, which sparked what has been termed the "Great Recession" in 2007. The findings of structural breaks in the correlation structure of the yield curves during these periods corroborate the findings of [44].

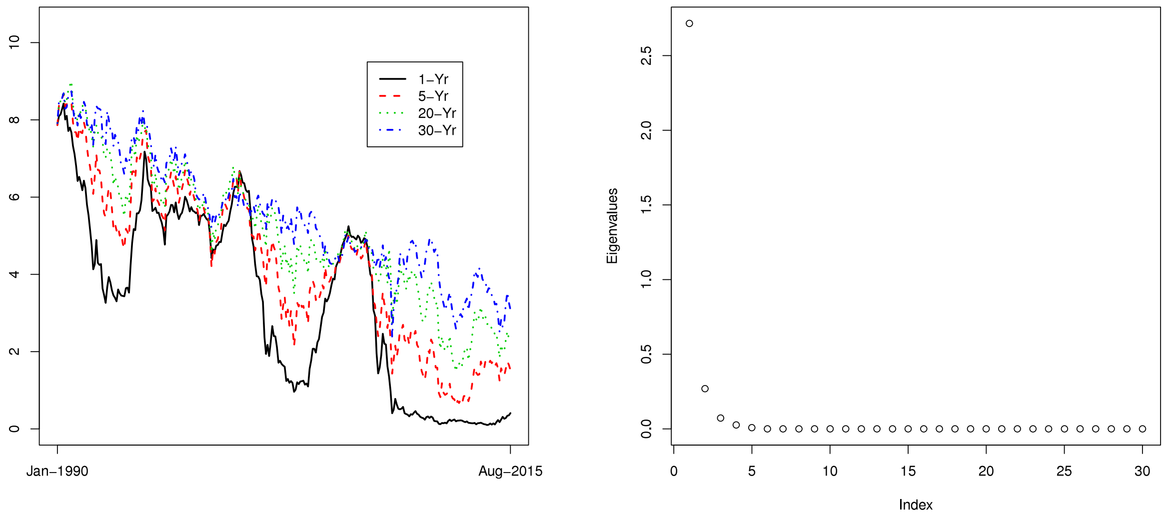

Fig. 4. (Left panel) Yield curves at a 1-month resolution between January, 1990 and August, 2015 corresponding to 1 year, 5 year, 10 year, and 30 year maturities. (Right panel) Screen plot of the eigenvalues of $\hat{\mathbf{C}}_{30,356}(1)$ calculated from the first differenced data. $\hat{\lambda}_1(1)/\mathrm{tr}\{\hat{\mathbf{C}}_{30,356}(1)\} \approx 0.88$.

Fig. 5. p-values corresponding to tests of $\mathcal{H}_0$ applied to 5-year blocks of the first differenced yield curve data. The vertical axis measures the magnitude of the p-value, and the horizontal axis indicates the concluding date of the 5-year block; p-values below the horizontal red line are below 0.05.

6. Results for smaller eigenvalues

In this section, we provide analogous results to Theorems 1 and 2 for the smaller eigenvalues. Namely, we aim to establish the weak convergence of the K-dimensional process

$$\mathbf{A}_{N,T}(u) = (A_{N,T,1}(u), \ldots, A_{N,T,K}(u))^\top,$$

where for $i \in \{1, \ldots, K\}$, $A_{N,T,i}(u) = T^{1/2}u\{\hat{\lambda}_i(u) - \lambda_i\}$ for u $\in [1/T, 1]$ and $A_{N,T,i}(u) = 0$ for $u \in [0, 1/T)$. Let

$$\sigma_\eta^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{cov}(\eta_0^2, \eta_\ell^2),$$

$$\mathbf{V}_2 = \left\{ \sum_{s=-\infty}^{\infty} \lim_{T\to\infty} \sum_{k=1}^{N} \mathfrak{e}_i(k)\mathfrak{e}_j(k)\mathrm{cov}(\eta_0, \eta_s)\mathrm{cov}(e_{k,0}, e_{k,s}) : 1 \le i, j \le K \right\},$$

$$\mathbf{V}_3 = \left\{ \sum_{s=-\infty}^{\infty} \lim_{T\to\infty} \left[\sum_{k=1}^{N} \mathfrak{e}_i^2(k)\mathfrak{e}_j^2(k)\mathrm{cov}(e_{k,0}^2, e_{k,s}^2) + 2\left\{ \sum_{k=1}^{N} \mathfrak{e}_i(k)\mathfrak{e}_j(k)\mathrm{cov}(e_{k,0}, e_{k,s}) \right\}^2 - 2\sum_{k=1}^{N} \mathfrak{e}_i^2(k)\mathfrak{e}_j^2(k)\{\mathrm{cov}(e_{k,0}, e_{k,s})\}^2 \right] : 1 \le i, j \le K \right\}.$$

We use the notation $\mathbf{V}_2 = \{V_2(i,j) : 1 \le i,j \le K\}$ and $\mathbf{V}_3 = \{V_3(i,j) : i,j \in \{1,\dots,K\}\}$.

Remark 5. If, for example, we assume that $r(s) = \mathrm{cov}(e_{k,0}, e_{k,s})$ for all $k \in \{1,\dots,N\}$, then $\mathbf{V}_2$ is a diagonal matrix with

$$V_2(i,i) = \sum_{s=-\infty}^{\infty} \mathrm{cov}(\eta_0, \eta_s) r(s).$$

The expression for $\mathbf{V}_3$ also simplifies since by the orthonormality of the $\mathfrak{e}_i$s we have

$$\sum_{k=1}^{N} \mathfrak{e}_i(k)\mathfrak{e}_j(k)\mathrm{cov}(e_{k,0}, e_{k,s}) = r(s)\mathbf{1}(i=j).$$

If we further assume that each of the $\{e_{k,s} : -\infty < s < \infty\}$ sequences are Gaussian, then $\mathrm{cov}(e_{k,0}^2, e_{k,s}^2) = 2r^2(s)$, and $\mathbf{V}_3$ also reduces to a diagonal matrix with $V_3(i,i) = 2\sum_{s=-\infty}^{\infty} r^2(s)$.

For each $i \in \{1,\dots,K\}$, let

$$a_i = \lim_{\min(N,T)\to\infty} \mathfrak{e}_i^\top \boldsymbol{\gamma}, \tag{18}$$

and define $\mathbf{G} = \{G(i,j) : 1 \le i,j \le K\}$ with $G(i,j) = a_i^2 a_j^2 \sigma_\eta^2 + 4a_i a_j V_2(i,j) + V_3(i,j)$. Lemma 4 demonstrates that the limit in (18) is finite.

Theorem 8. *Suppose $\mathcal{H}_0$ and Assumptions 1–3, (6) hold, and $\|\boldsymbol{\gamma}\| = O(1)$ as $\min(N,T) \to \infty$, then we have that $\mathbf{A}_{N,T}(u)$ converges weakly in $\mathcal{D}^K[0,1]$ to $\mathbf{W}_{\mathbf{G}}(u)$, where $\mathbf{W}_{\mathbf{G}}(u)$ is a K-dimensional Wiener process, i.e., $\mathbf{W}_{\mathbf{G}}(u)$ is Gaussian with $\mathrm{E}\mathbf{W}_{\mathbf{G}}(u) = \mathbf{0}$ and $\mathrm{E}\mathbf{W}_{\mathbf{G}}(u)\mathbf{W}_{\mathbf{G}}^\top(u') = \min(u,u')\mathbf{G}$.*

Remark 6. If $\|\boldsymbol{\gamma}\| \to 0$, as $\min(N,T) \to \infty$, then $a_i = 0$ according to Lemma 4. In this case the weak limit of $\mathbf{A}_{N,T}(u)$ is the K-dimensional Wiener process $\mathbf{W}_{\mathbf{V}_3}(u)$, since $\mathbf{G} = \mathbf{V}_3$.

To state the next result we introduce the covariance matrix $\mathbf{H} = \{H(i,j) : 1 \le i,j \le K\}$ by setting $H(1,1) = \sigma_\eta^2$, $H(1,i) = H(i,1) = a_i^2\sigma_\eta^2$ and $H(i,j) = a_i^2 a_j^2 \sigma_\eta^2 + 4a_i a_j V_2(i,j) + V_3(i,j)$ for all $i,j \in \{2,\dots,K\}$.

Theorem 9. *Suppose $\mathcal{H}_0$ and Assumptions 1–3, (6) hold, and $\|\boldsymbol{\gamma}\| \to \infty$ as $\min(N,T) \to \infty$, then we have that $\{\|\boldsymbol{\gamma}\|^{-2}A_{N,T;1}(u), A_{N,T;i}(u), 2 \le i \le K\}$ converges weakly in $\mathcal{D}^K[0,1]$ to $\mathbf{W}_{\mathbf{H}}(u)$, where $\mathbf{W}_{\mathbf{H}}(u)$ is a K-dimensional Wiener process, i.e., $\mathbf{W}_{\mathbf{H}}(u)$ is Gaussian with $\mathrm{E}\mathbf{W}_{\mathbf{H}}(u) = \mathbf{0}$ and $\mathrm{E}\{\mathbf{W}_{\mathbf{H}}(u)\mathbf{W}_{\mathbf{H}}^\top(u')\} = \min(u,u')\mathbf{H}$.*

Remark 7. We note that σ_1^2 defined in Theorem 2 coincide with $G(1,1)$ and $H(1,1)\|\gamma\|^2$ in the cases when $\|\boldsymbol{\gamma}\| = O(1)$ and $\|\boldsymbol{\gamma}\| \to \infty$ as $\min(N,T) \to \infty$, respectively.

Remark 8. Theorems 2, 8 and 9 provide the limits of the weighted differences $T^{1/2}u\{\hat{\lambda}_i(u) - \lambda_i\} = T^{1/2}(\tilde{\lambda}_i - u\lambda_i)$ for all $i \in \{1,\dots,K\}$. If the conditions of Theorem 1 are satisfied and (4) holds, then for each $i \in \{1,\dots,K\}$, $T^{1/2}\{\hat{\lambda}_i(u) - \lambda_i\}$ converges weakly in $\mathcal{D}^K[c,1]$ to $\mathbf{W}_{\mathbf{G}}(u)/u$ for any $c \in (0,1]$ where $\mathbf{W}_{\mathbf{G}}(u)$ is defined in Theorem 2.

6.1. *Testing using first d largest eigenvalues*

Theorems 8 and 9 may be used to construct tests $\mathcal{H}_0$ based on the first d largest eigenvalue. Let for $u \in (0,1)$

$$\hat{B}_{T,j}(u) = T^{1/2}\, u\{\hat{\lambda}_j(u) - \hat{\lambda}_j(1)\}/\hat{v}_{j,T},$$

where $\hat{v}_{j,T}^2$ is defined as in (9). One can show that under the conditions of Theorem 4, $\hat{v}_{j,T}^2$ converges to either $\mathbf{H}(j,j)$ or $\mathbf{G}(j,j)$ if $\|\boldsymbol{\gamma}\| \to \infty$ or $\|\boldsymbol{\gamma}\| = O(1)$, respectively. Therefore, one can perform a conservative test of size α for $\mathcal{H}_0$ based on d eigenvalues by comparing $\max_{j\in\{1,\dots,d\}} \sup_{0\le u\le 1} |\hat{B}_{T,j}(u)|$, to the α/d critical value of the Kolmogorov distribution, where the choice α/d is motivated by Bonferroni's inequality.

7. Conclusion

We considered the problem of testing for generic structural breaks in high-dimensional linear factor models by way of analyzing the processes of partial sample estimates of the largest eigenvalues of the covariance matrix. We showed that a Gaussian approximation may be obtained for these processes in two cases: one is when the contribution of the common factor to the covariance structure of the cross sections is dominate, and in this case we only require mild conditions on the divergence rates of N and T for the approximation to hold. Secondly, in the case when the common factor is relatively negligible, we showed that a Gaussian approximation for these processes is only attainable when $N/T^{1/2} \to 0$. These results can be used to test for the stability of factor models by considering sup functionals of a CUSUM processes derived from the eigenvalue processes. As it is unknown in practice whether or not there is a dominant common factor, we derived normalizing sequence estimators that converge in either situation, as well as ratio type test statistics. We studied the theoretical and empirical properties of such test statistics under both $\mathcal{H}_0$ and $\mathcal{H}_A$. Finally, we applied these results to US treasury yield data, and the results demonstrate that there was a pronounced structural break in the largest eigenvalue of the covariance matrix following the US subprime crisis.

8. Technical results

8.1. Proof of Theorems 2–3, 8 and 9

Throughout these proofs we use the terms of the form $c_{i,j}$ to denote unimportant numerical constants. We can assume without loss of generality that $E\mathbf{X}_t = \mathbf{0}$, and so we define

$$\mathbf{C}_{N,T}(u) = \frac{1}{T}\sum_{t=1}^{\lfloor Tu \rfloor} \mathbf{X}_t \mathbf{X}_t^\top .$$

Lemma 1. *If (1) and Assumptions 1–3 hold, then we have, as* $\min(N,T) \to \infty$,

$$\sup_{u\in[0,1]} \|\tilde{\mathbf{C}}_{N,T}(u) - \mathbf{C}_{N,T}(u)\| = O_P\left(N/T\right).$$

Proof. It is easy to see that

$$\tilde{\mathbf{C}}_{N,T}(u) = \mathbf{C}_{N,T}(u) - \bar{\mathbf{X}}_T\left(\frac{1}{T}\sum_{t=1}^{\lfloor Tu \rfloor}\mathbf{X}_t\right)^\top - \left(\frac{1}{T}\sum_{t=1}^{\lfloor Tu \rfloor}\mathbf{X}_t\right)\bar{\mathbf{X}}_T^\top + \bar{\mathbf{X}}_T\bar{\mathbf{X}}_T^\top,$$

and therefore

$$\sup_{u\in[0,1]} \|\tilde{\mathbf{C}}_{N,T}(u) - \mathbf{C}_{N,T}(u)\| \le 2\sup_{u\in[0,1]}\left\|\bar{\mathbf{X}}_T\left(\frac{1}{T}\sum_{t=1}^{\lfloor Tu \rfloor}\mathbf{X}_t\right)^\top\right\| + \left\|\bar{\mathbf{X}}_T\bar{\mathbf{X}}_T^\top\right\| \le 3\sup_{u\in[0,1]}\left\|\bar{\mathbf{X}}_T\left(\frac{1}{T}\sum_{t=1}^{\lfloor Tu \rfloor}\mathbf{X}_t\right)^\top\right\|.$$

Using model (1) we obtain that

$$\begin{aligned}
T^2\left\|\sum_{t=1}^{\lfloor Tu \rfloor}\mathbf{X}_t\bar{\mathbf{X}}_T^\top\right\|^2 &= \sum_{\ell=1}^{N}\sum_{p=1}^{N}\left(\gamma_\ell\sum_{t=1}^{T}\eta_t + \sum_{t=1}^{T}e_{\ell,t}\right)^2\left(\gamma_p\sum_{t=1}^{\lfloor Tu \rfloor}\eta_t + \sum_{t=1}^{\lfloor Tu \rfloor}e_{p,t}\right)^2\\
&= \left(\sum_{\ell=1}^{N}\gamma_\ell^2\right)^2\left(\sum_{t=1}^{\lfloor Tu \rfloor}\eta_t\right)^2\left(\sum_{t=1}^{T}\eta_t\right)^2 + 2\sum_{\ell=1}^{N}\gamma_\ell^2\left(\sum_{t=1}^{\top}\eta_t\right)^2\left(\sum_{v=1}^{\lfloor Tu \rfloor}\eta_v\right)\left(\sum_{s=1}^{\lfloor Tu \rfloor}\sum_{p=1}^{N}\gamma_p e_{p,s}\right)\\
&\quad + \sum_{\ell=1}^{N}\gamma_\ell^2\left(\sum_{t=1}^{\top}\eta_t\right)^2\sum_{p=1}^{N}\left(\sum_{t=1}^{\lfloor Tu \rfloor}e_{p,t}\right)^2 + \sum_{p=1}^{N}\gamma_p^2\left(\sum_{t=1}^{\lfloor Tu \rfloor}\eta_t\right)^2\sum_{\ell=1}^{N}\left(\sum_{s=1}^{\top}e_{\ell,s}\right)^2\\
&\quad + 2\sum_{\ell=1}^{N}\left(\sum_{s=1}^{\top}e_{\ell,s}\right)^2\left(\sum_{s=1}^{\lfloor Tu \rfloor}\eta_s\right)\left(\sum_{s=1}^{\lfloor Tu \rfloor}\sum_{p=1}^{N}\gamma_p e_{p,s}\right) + \sum_{\ell=1}^{N}\left(\sum_{s=1}^{\top}e_{\ell,s}\right)^2\sum_{p=1}^{N}\left(\sum_{s=1}^{\lfloor Tu \rfloor}e_{p,s}\right)^2\\
&\quad + 2\sum_{t=1}^{\top}\eta_t\sum_{\ell=1}^{N}\gamma_\ell\left(\sum_{s=1}^{\top}e_{\ell,s}\right)\sum_{p=1}^{N}\gamma_p^2\left(\sum_{v=1}^{\lfloor Tu \rfloor}\eta_v\right)^2 + 2\sum_{t=1}^{\top}\eta_t\sum_{\ell=1}^{N}\gamma_\ell\left(\sum_{s=1}^{\top}e_{\ell,s}\right)\sum_{p=1}^{N}\left(\sum_{s=1}^{\lfloor Tu \rfloor}e_{p,s}\right)^2
\end{aligned}$$

$$+4\sum_{t=1}^{\top}\eta_t\sum_{\ell=1}^{N}\left(\sum_{s=1}^{\top}\gamma_\ell e_{\ell,s}\right)\sum_{z=1}^{\lfloor Tu\rfloor}\eta_z\sum_{p=1}^{N}\left(\sum_{s=1}^{\lfloor Tu\rfloor}\gamma_p e_{p,s}\right)$$

$$\equiv R_{T,1}(u)+\cdots+R_{T,9}(u).$$

First we prove that

$$\sup_{u\in[0,1]}\left|\sum_{t=1}^{\lfloor Tu\rfloor}\eta_t\right| = O_P(T^{1/2}). \tag{19}$$

It follows from Proposition 4 of [7] that under conditions Assumption 3(a) and Assumption 3(a) we have for any $\kappa\in(2,12]$ that, for all $u, v\in[0,1]$ with $v\le u$

$$\mathrm{E}\left(\sum_{t=\lfloor Tv\rfloor}^{\lfloor Tu\rfloor}\eta_t\right)^{\kappa} \le c_{1,1}(\lfloor Tu\rfloor-\lfloor Tv\rfloor)^{\kappa/2}, \tag{20}$$

and therefore the maximal inequality of [31] implies (19). Next we show that

$$\sup_{u\in[0,1]}\left|\sum_{s=1}^{\lfloor Tu\rfloor}\sum_{p=1}^{N}\gamma_p e_{p,s}\right| = O_P(1)T^{1/2}\|\boldsymbol{\gamma}\|. \tag{21}$$

Following the arguments leading to (20) one can verify that for any $\kappa\in(2,12]$ and all $u, v\in[0,1]$ with $v\le u$

$$\mathrm{E}\left|\sum_{s=\lfloor Tv\rfloor}^{\lfloor Tu\rfloor}e_{p,s}\right|^{\kappa} \le c_{1,2}(\lfloor Tu\rfloor-\lfloor Tv\rfloor)^{\kappa/2}, \tag{22}$$

with some constant $c_{1,2}$ for all $p\in\{1,\ldots,N\}$. Hence for any $0\le v<u\le 1$ we have via Rosenthal's inequality ([35], p. 59) and (22) that

$$\mathrm{E}\left|\sum_{s=\lfloor Tv\rfloor}^{\lfloor Tu\rfloor}\sum_{p=1}^{N}\gamma_p e_{p,s}\right|^{\kappa} = \mathrm{E}\left|\sum_{p=1}^{N}\sum_{s=\lfloor Tv\rfloor}^{\lfloor Tu\rfloor}\gamma_p e_{p,s}\right|^{\kappa} \le c_{1,3}\left\{\sum_{p=1}^{N}|\gamma_p|^{\kappa}\mathrm{E}\left|\sum_{s=\lfloor Tv\rfloor}^{\lfloor Tu\rfloor}e_{p,s}\right|^{\kappa}+\left(\sum_{p=1}^{N}\gamma_p^2\mathrm{E}\left(\sum_{s=\lfloor Tv\rfloor}^{\lfloor Tu\rfloor}e_{p,s}\right)^2\right)^{\kappa/2}\right\}$$

$$\le c_{1,4}(\lfloor Tu\rfloor-\lfloor Tv\rfloor)^{\kappa/2}\left\{\sum_{p=1}^{N}|\gamma_p|^{\kappa}+\left(\sum_{p=1}^{N}\gamma_p^2\right)^{\kappa/2}\right\}.$$

Using again the maximal inequality of [31] we conclude

$$\mathrm{E}\sup_{u\in[0,1]}\left|\sum_{s=1}^{\lfloor Tu\rfloor}\sum_{p=1}^{N}\gamma_p e_{s,p}\right|^{\kappa} \le c_{1,5}T^{\kappa/2}\left\{\sum_{p=1}^{N}|\gamma_p|^{\kappa}+\left(\sum_{p=1}^{N}\gamma_p^2\right)^{\kappa/2}\right\}\le c_{1,6}T^{\kappa/2}\|\boldsymbol{\gamma}\|^{\kappa},$$

by Assumption 2. This completes the proof of (21). Similarly to (21) we show that

$$\sup_{0\le s\le 1}\sum_{\ell=1}^{N}\left(\sum_{s=1}^{\lfloor Tu\rfloor}e_{\ell,s}\right)^2 = O_P(NT). \tag{23}$$

First we note

$$\mathrm{E}\sup_{u\in[0,1]}\sum_{\ell=1}^{N}\left(\sum_{s=1}^{\lfloor Tu\rfloor}e_{\ell,s}\right)^2 \le \sum_{\ell=1}^{N}\mathrm{E}\sup_{u\in[0,1]}\left(\sum_{s=1}^{\lfloor Tu\rfloor}e_{\ell,s}\right)^2$$

and by Jensen's inequality we have

$$\mathrm{E}\sup_{u\in[0,1]}\left(\sum_{s=1}^{\lfloor Tu\rfloor}e_{\ell,s}\right)^2 \le \left(\mathrm{E}\sup_{u\in[0,1]}\left|\sum_{s=1}^{\lfloor Tu\rfloor}e_{\ell,s}\right|^{\kappa}\right)^{2/\kappa}.$$

Using again Proposition 4 of [7] we get for all $u, v\in[0,1]$ with $v\le u$ that

$$\mathrm{E}\left|\sum_{s=\lfloor Tv\rfloor}^{\lfloor Tu\rfloor}e_{\ell,s}\right|^{\kappa} \le c_{1,7}(\lfloor Tu\rfloor-\lfloor Tv\rfloor)^{\kappa/2}$$

and therefore the maximal inequality of [31] yields

$$\left(\mathrm{E}\sup_{u\in[0,1]}\left|\sum_{s=1}^{\lfloor Tu\rfloor}e_{\ell,s}\right|^{\kappa}\right)^{2/\kappa}\leq c_{1,8}T^{1/2}.$$

This completes the proof of (21). The upper bounds in (19)–(23) imply

$$\sup_{u\in[0,1]}|R_{T,i}(u)|=O_P((\|\boldsymbol{\gamma}\|^4+\|\boldsymbol{\gamma}\|^3)T^2),\quad \text{if } i\in\{1,2,7\},\quad \sup_{u\in[0,1]}|R_{T,i}(u)|=O_P((\|\boldsymbol{\gamma}\|^2+\|\boldsymbol{\gamma}\|)NT^2),\quad \text{if } i\in\{3,4,5,8,9\}.$$

and

$$\sup_{u\in[0,1]}|R_{T,6}(u)|=O_P(N^2T^2).$$

Assumption 2 implies that $\|\boldsymbol{\gamma}\|\leq c_{1,9}N$, the proof of Lemma 1 is complete. □

Let $\bar\lambda_1(u)\geq\cdots\geq\bar\lambda_K(u)$ denote the K largest eigenvalues of $\mathbf{C}_{N,T}(u)$.

Lemma 2. *If (1) and Assumptions 1–3 hold, then we have, as* $\min(N,T)\to\infty$,

$$\max_{i\in\{1,\ldots,K\}}\sup_{u\in[0,1]}|\tilde\lambda_i(u)-\bar\lambda_i(u)|=O_P\left(N/T\right).$$

Proof. It is well-known (see, e.g., p. 577 of [11]) that

$$\max_{i\in\{1,\ldots,K\}}\sup_{u\in[0,1]}|\tilde\lambda_i(u)-\bar\lambda_i(u)|\leq\sup_{u\in[0,1]}\|\tilde{\mathbf{C}}_{N,T}(u)-\mathbf{C}_{N,T}(u)\|,$$

and therefore the result follows from Lemma 1. □

For each $i\in\{1,\ldots,K\}$, let

$$Z_{N,T;i}(u)=\sum_{\ell\neq i}^{N}\frac{1}{u(\lambda_i-\lambda_\ell)}[\mathfrak{e}_i^\top\{\mathbf{C}_{N,T}(u)-u\mathbf{C}\}\mathfrak{e}_\ell]^2.$$

Lemma 3. *If (1), Assumptions 1–3 hold, then we have, as* $\min(N,T)\to\infty$,

$$\sup_{u\in[0,1]}|\bar\lambda_i(u)-(\lfloor Tu\rfloor/T)\lambda_i-\mathfrak{e}_i^\top\{\mathbf{C}_{N,T}(u)-u\mathbf{C}\}\mathfrak{e}_i-Z_{N,T;i}(u)|=O_P(N^2T^{-3/2}).$$

Proof. According to formula (5.17) of [15] we have

$$\sup_{u\in[0,1]}|\bar\lambda_i(u)-(\lfloor Tu\rfloor/T)\lambda_i-\mathfrak{e}_i^\top\{\mathbf{C}_{N,T}(u)-u\mathbf{C}\}\mathfrak{e}_i-Z_{N,T;i}(u)|\leq c_{2,1}\sup_{u\in[0,1]}\Delta^3(u),$$

where $\Delta(u)=\max_{\ell\in\{1,\ldots,N\}}R_{N,T;\ell}(u)$ with

$$R_{N,T;\ell}(u)=\left[\sum_{j=1}^{N}\{C_{N,T;j,\ell}(u)-(\lfloor Tu\rfloor/T)C_{j,\ell}\}^2\right]^{1/2},$$

and $C_{N,T;j,\ell}(u)$ and $C_{j,\ell}$ denote the (j,ℓ)th element of $\hat{\mathbf{C}}_{N,T}(u)$ and $\mathbf{C}$, respectively. By inequality (2.30) in [35], p. 58, we conclude

$$R^6_{N,T;\ell}(u)\leq N^2\sum_{j=1}^{N}\{C_{N,T;j,\ell}(u)-(\lfloor Tu\rfloor/T)C_{j,\ell}\}^6$$

and hence

$$\mathrm{E}\sup_{u\in[0,1]}\{R_{N,T;\ell}(u)\}^6\leq N^2\sum_{j=1}^{N}\mathrm{E}\sup_{u\in[0,1]}\{C_{N,T;j,\ell}(u)-(\lfloor Tu\rfloor/T)C_{j,\ell}\}^6.$$

Using the definitions of $C_{N,T;j,\ell}(u)$ and $C_{j,\ell}$, we write

$$\left\{C_{N,T;j,\ell}(u)-\frac{\lfloor Tu\rfloor}{T}C_{j,\ell}\right\}^6=T^{-6}\left[\sum_{s=1}^{\lfloor Tu\rfloor}\{\gamma_\ell\gamma_j(\eta_s^2-1)+\gamma_\ell\eta_s e_{j,s}+\gamma_j\eta_s e_{\ell,s}+e_{\ell,s}e_{j,s}-\mathrm{E}e_{\ell,s}e_{j,s}\}\right]^6$$
$$\le 4^6T^{-6}\left[\gamma_\ell^6\gamma_j^6\left\{\sum_{s=1}^{\lfloor Tu\rfloor}(\eta_s^2-1)\right\}^6+\gamma_\ell^6\left(\sum_{s=1}^{\lfloor Tu\rfloor}\eta_s e_{j,s}\right)^6+\gamma_j^6\left(\sum_{s=1}^{\lfloor Tu\rfloor}\eta_s e_{\ell,s}\right)^6\right.$$
$$\left.+\left\{\sum_{s=1}^{\lfloor Tu\rfloor}(e_{\ell,s}e_{j,s}-\mathrm{E}e_{\ell,s}e_{j,s})\right\}^6\right].$$

Utilizing Assumption 3(a), we obtain along the lines of (20) that $\mathrm{E}\{\sum_{s=1}^{\top}(\eta_s^2-1)\}^6\le c_{2,2}t^3$, so by the stationarity of $\{\eta_t^2:-\infty<t<\infty\}$ and the maximal inequality of [31] we obtain that

$$\mathrm{E}\sup_{u\in[0,1]}\left\{\sum_{s=1}^{\lfloor Tu\rfloor}(\eta_s^2-1)\right\}^6\le c_{2,3}T^3.$$

Similarly, for all $j,\ell\in\{1,\dots,N\}$,

$$\mathrm{E}\sup_{u\in[0,1]}\left(\sum_{s=1}^{\lfloor Tu\rfloor}\eta_s e_{\ell,s}\right)^6\le c_{2,4}T^3\quad\text{and}\quad \mathrm{E}\sup_{u\in[0,1]}\left\{\sum_{s=1}^{\lfloor Tu\rfloor}(e_{\ell,s}e_{j,s}-\mathrm{E}e_{\ell,s}e_{j,s})\right\}^6\le c_{2,5}T^3.$$

Hence for all $\ell\in\{1,\dots,N\}$, we have by Assumption 2 that

$$\mathrm{E}\left\{\sup_{u\in[0,1]}R_{N,T;\ell}(u)\right\}^6\le c_{2,6}T^{-3}N^3. \tag{24}$$

Using (24) we conclude that, for all $x>0$,

$$\Pr\left\{\sup_{u\in[0,1]}\max_{\ell\in\{1,\dots,N\}}R_{N,T;\ell}(u)>xN^{2/3}T^{-1/2}\right\}\le\sum_{\ell=1}^N\Pr\left\{\sup_{u\in[0,1]}R_{N,T;\ell}(u)>xN^{2/3}T^{-1/2}\right\}$$
$$\le\sum_{\ell=1}^N\frac{T^3}{x^6N^6}\mathrm{E}\left\{\sup_{u\in[0,1]}R_{N,T;\ell}(u)\right\}^6\le\sum_{\ell=1}^N\frac{T^3}{x^6N^6}C_5T^{-3}N^3,$$

which shows that

$$\sup_{u\in[0,1]}\Delta^3(u)=O_P(N^2T^{-3/2})\quad\text{and}\quad\sup_{u\in[0,1]}u\hat{\Delta}^3(u)=O_P(N^2T^{-3/2}).\quad\square$$

Since $\mathfrak{e}_1$ is defined via (2) up to a sign, we can assume without loss of generality that $\boldsymbol{\gamma}^\top\mathfrak{e}_1\ge 0$.

Lemma 4. *If Assumptions 1–3 hold and $\|\boldsymbol{\gamma}\|\to\infty$ hold, then we have*

$$\|\mathfrak{e}_1-\boldsymbol{\gamma}/\|\boldsymbol{\gamma}\|\|=O(1/\|\boldsymbol{\gamma}\|), \tag{25}$$

$$\lambda_1/\|\boldsymbol{\gamma}\|^2\to 1, \tag{26}$$

$$\max(|\boldsymbol{\gamma}^\top\mathfrak{e}_2|,\dots,|\boldsymbol{\gamma}^\top\mathfrak{e}_N|)\le c_{3,1}\ \ \textit{with some constant}\ \ c_{3,1}, \tag{27}$$

and

$$\max(\lambda_2,\dots,\lambda_N)\le c_{3,2}\ \ \textit{with some constant}\ \ c_{3,2}.$$

Proof. By (1) we have $\mathbf{C}=\boldsymbol{\gamma}\boldsymbol{\gamma}^\top+\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is the $N\times N$ diagonal matrix with $\sigma_1^2,\dots,\sigma_N^2$ in the diagonal. We can write

$$\mathfrak{e}_1=\bar{\alpha}_1\frac{\boldsymbol{\gamma}}{\|\boldsymbol{\gamma}\|}+\bar{\beta}_1\mathbf{r}_1,\ \text{with some}\ \bar{\alpha}_1\ge 0,\ \text{where}\ \bar{\alpha}_1^2+\bar{\beta}_1^2=1,\ \boldsymbol{\gamma}^\top\mathbf{r}_1=0\ \text{and}\ \|\mathbf{r}_1\|=1.$$

It follows from the definition of λ_1 and $\mathfrak{e}_1$ that $\lambda_1=\mathfrak{e}_1^\top\mathbf{C}\mathfrak{e}_1\ge\|\boldsymbol{\gamma}\|^2$, and

$$\mathfrak{e}_1^\top\mathbf{C}\mathfrak{e}_1=\bar{\alpha}_1^2\|\boldsymbol{\gamma}\|^2+\mathfrak{e}_1^\top\boldsymbol{\Lambda}\mathbf{C}\mathfrak{e}_1,\quad\mathfrak{e}_1^\top\boldsymbol{\Lambda}\mathbf{C}\mathfrak{e}_1\le\sum_{\ell=1}^N\mathfrak{e}_i^2(\ell)\sigma_\ell^2\le c_5 \tag{28}$$

where c_5 is defined in Assumption 3(b). Thus we conclude $\|\boldsymbol{\gamma}\|^2 \le \bar{\alpha}_1^2\|\boldsymbol{\gamma}\|^2 + c_5$. By assumption $\boldsymbol{\gamma}^\top \mathfrak{e}_1 \ge 0$ and therefore $\bar{\alpha}_1 \in [0, 1]$. Hence $(1-\bar{\alpha}_1)^2 \le 1-\bar{\alpha}_1^2 \le c_5/\|\boldsymbol{\gamma}\|^2$ and $\bar{\beta}_1^2 \le c_5/\|\boldsymbol{\gamma}\|^2$. Thus we get

$$\|\mathfrak{e}_1 - \boldsymbol{\gamma}/\|\boldsymbol{\gamma}\|\|^2 = (1-\bar{\alpha}_1)^2 + \bar{\beta}_1^2 \le 2c_5/\|\boldsymbol{\gamma}\|^2, \tag{29}$$

completing the proof of (25). Since $\bar{\alpha}_2^2 \ge 1 - c_5/\|\boldsymbol{\gamma}\|^2$, (26) follows from (28). For all $i \in \{2, 3, \ldots\}$ we have

$$|\boldsymbol{\gamma}^\top \mathfrak{e}_i| = \|\boldsymbol{\gamma}\| \left|(\boldsymbol{\gamma}/\|\boldsymbol{\gamma}\| - \mathfrak{e}_1)^\top \mathfrak{e}_i\right| \le \|\boldsymbol{\gamma}\| \, \|\boldsymbol{\gamma}/\|\boldsymbol{\gamma}\| - \mathfrak{e}_1\| \le 2c_5$$

by (29) which gives (27). Since $\lambda_i = \mathfrak{e}_i^\top \mathbf{C}\mathfrak{e}_i = (\mathfrak{e}_i^\top\boldsymbol{\gamma})^2 + \mathfrak{e}_i^\top \boldsymbol{\Lambda}\mathfrak{e}_i$ and $\mathfrak{e}_i^\top \boldsymbol{\Lambda}\mathfrak{e}_i = \sum_{\ell=1}^N \mathfrak{e}_i^2(\ell)\sigma_\ell^2 \le c_5$ by Assumption 3(b), the last claim of this lemma follows from (27). □

Lemma 5. *If (1), and Assumptions 1–3 hold, then we have*

$$\max_{i\in\{1,\ldots,K\}} \sup_{u\in[0,1]} |Z_{N,T;i}(u)| = O_P\{N(\ln T)^{1/3}/T\}. \tag{30}$$

Proof. It follows from (2) that $\mathfrak{e}_i\mathbf{C}\mathfrak{e}_\ell = 0$, if $i \ne \ell$. Hence we get

$$\mathfrak{e}_i^\top\{\mathbf{C}_{N,T}(u) - (\lfloor Tu\rfloor/T)\mathbf{C}\}\mathfrak{e}_\ell = \frac{1}{T}\sum_{s=1}^{\lfloor Tu\rfloor} \mathfrak{e}_i^\top \mathbf{X}_s\mathbf{X}_s^\top \mathfrak{e}_\ell.$$

First we assume that $\|\boldsymbol{\gamma}\| = O(1)$. It follows from the definition of $Z_{N,T;i}$ that

$$|Z_{N,T;i}(u)| = \left|\sum_{\ell\ne i}^N \frac{1}{u(\lambda_i-\lambda_\ell)}\{\mathfrak{e}_i^\top(\mathbf{C}_{N,T}(u) - u\mathbf{C})\mathfrak{e}_\ell\}^2\right| \le \frac{1}{c_5}\frac{1}{T}\sum_{\ell\ne i}^N\left\{\frac{1}{(Tu)^{1/2}}\sum_{s=1}^{\lfloor Tu\rfloor}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right\}^2,$$

where c_0 is defined in Assumption 1. Let $\rho > 1$ and write with $c = \lfloor 1/\ln\rho\rfloor + 1$

$$\max_{v\in\{1,\ldots,T\}} v^{-1/2}\left|\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right| \le \max_{1\le k\le c\ln T}\max_{\rho^{k-1}<v\le\rho^k} v^{-1/2}\left|\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right| \le \max_{1\le k\le c\ln T}\rho^{-(k-1)/2}\max_{1\le v\le\rho^k}\left|\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right|.$$

Thus we get for any $x > 0$ via Markov's inequality that

$$\begin{aligned}\Pr\left\{\max_{v\in\{1,\ldots,T\}} v^{-1/2}\left|\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right| > x\right\} &\le \sum_{k=1}^{c\ln T}\Pr\left\{\max_{1\le v\le\rho^k}\left|\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right| > x\rho^{(k-1)/2}\right\}\\ &\le \sum_{k=1}^{c\ln T} x^{-6}\rho^{-3(k-1)}\mathrm{E}\left(\max_{1\le v\le\rho^k}\left|\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right|\right)^6.\end{aligned} \tag{31}$$

Using (1) we obtain with $\mathfrak{e}_i = (\mathfrak{e}_i(1), \ldots, \mathfrak{e}_i(N))^\top$ that

$$\begin{aligned}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell &= \sum_{k=1}^N\gamma_k\mathfrak{e}_i(k)\sum_{n=1}^N\gamma_n\mathfrak{e}_\ell(n)(\eta_s^2-1) + \sum_{k=1}^N\gamma_k\mathfrak{e}_i(k)\eta_s\sum_{n=1}^N e_{n,s}\mathfrak{e}_\ell(n)\\ &\quad + \sum_{n=1}^N\gamma_n\mathfrak{e}_\ell(n)\eta_s\sum_{k=1}^N e_{k,s}\mathfrak{e}_i(k) + \sum_{n=1}^N\sum_{k=1}^N(e_{k,s}e_{n,s} - \mathrm{E}e_{k,s}e_{n,s})\mathfrak{e}_i(k)\mathfrak{e}_\ell(n),\end{aligned}$$

since for $i \ne \ell$ we have $\mathrm{E}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell = \mathfrak{e}_i^\top\mathbf{C}\mathfrak{e}_\ell = 0$. Clearly, on account of $\|\mathfrak{e}_i\| = 1$, the Cauchy–Schwarz inequality implies

$$\left|\sum_{k=1}^N\gamma_k\mathfrak{e}_i(k)\right| \le \|\boldsymbol{\gamma}\|.$$

Following the proofs of (20), we get that from Assumption 3(a) that

$$\mathrm{E}\left|\sum_{k=1}^N\gamma_k\mathfrak{e}_i(k)\sum_{m=1}^N\gamma_m\mathfrak{e}_m(\ell)\sum_{s=1}^{v}(\eta_s^2-1)\right|^6 \le c_{4,1}v^3\|\boldsymbol{\gamma}\|^{12}. \tag{32}$$

Let

$$\tau_s = \eta_s\sum_{n=1}^N e_{n,s}\mathfrak{e}_\ell(n) \quad \text{and} \quad \tau_s^{(m)} = \eta_s^{(m)}\sum_{n=1}^N e_{n,s}^{(m)}\mathfrak{e}_\ell(n),$$

where $\eta_s^{(m)}$ and $e_{n,s}^{(m)}$ are defined in Assumption 3(a) and Assumption 3(b), respectively. By independence we have

$$E|\tau_0-\tau_0^{(m)}|^6 \le 2^6 E|\eta_0-\eta_0^{(m)}|^6 E\left|\sum_{n=1}^N e_{n,0}\mathfrak{e}_\ell(n)\right|^6 + 2^6 E|\eta_0^{(m)}|^6 E\left|\sum_{n=1}^N (e_{n,0}-e_{n,0}^{(m)})\mathfrak{e}_\ell(n)\right|^6.$$

By the independence of the variables $e_{1,0},\ldots,e_{N,0}$ and the Rosenthal inequality [35], we conclude that

$$E\left|\sum_{n=1}^N e_{n,0}\mathfrak{e}_\ell(n)\right|^6 \le c_{4,2}\left[\sum_{n=1}^N E|e_{n,0}|^6|\mathfrak{e}_\ell(n)|^6 + \left\{\sum_{n=1}^N Ee_{n,0}^2\mathfrak{e}_\ell^2(n)\right\}^3\right] \le c_{4,3}\sup_{1\le n<\infty} Ee_{n,0}^6 \le c_{4,4},$$

where $c_{4,4}$ is a constant, on account of Assumption 3(b) and $\|\mathfrak{e}_\ell\|=1$. Due to the independence of $e_{n,0}-e_{n,0}^{(m)}$ and $e_{r,0}-e_{r,0}^{(m)}$, if $n\ne r$, we can apply again the Rosenthal inequality to get

$$E\left|\sum_{n=1}^N (e_{n,0}-e_{n,0}^{(m)})\mathfrak{e}_\ell(n)\right|^6 \le c_{4,5}\left[\sum_{n=1}^N E|e_{n,0}-e_{n,0}^{(m)}|^6|\mathfrak{e}_\ell(n)|^6 + \left\{\sum_{n=1}^N E(e_{n,0}-e_{n,0}^{(m)})^2\mathfrak{e}_\ell^2(n)\right\}^3\right] \le c_{4,6}m^{-6\alpha},$$

resulting in

$$E|\tau_0-\tau_0^{(m)}|^6 \le c_{4,7}m^{-6\alpha}. \tag{33}$$

Hence the moment inequality in [7] yields

$$E\left|\sum_{s=1}^v \tau_s\right|^6 \le c_{4,8}v^3. \tag{34}$$

Similarly to (34) we have

$$E\left|\sum_{s=1}^v \eta_s\sum_{k=1}^N e_{k,s}\mathfrak{e}_i(k)\right|^6 \le c_{4,9}v^3. \tag{35}$$

Let

$$\bar\tau_s = \sum_{n=1}^N\sum_{k=1}^N (e_{k,s}e_{n,s}-Ee_{k,s}e_{n,s})\mathfrak{e}_i(k)\mathfrak{e}_\ell(n) = \sum_{n=1}^N e_{n,s}\mathfrak{e}_\ell(n)\sum_{k=1}^N e_{k,s}\mathfrak{e}_i(k) - \sum_{n=1}^N Ee_{n,s}^2\mathfrak{e}_i(n)\mathfrak{e}_\ell(n)$$

and

$$\bar\tau_s^{(m)} = \sum_{n=1}^N\sum_{k=1}^N (e_{k,s}^{(m)}e_{n,s}^{(m)}-Ee_{k,s}e_{n,s})\mathfrak{e}_i(k)\mathfrak{e}_\ell(n) = \sum_{n=1}^N e_{n,s}^{(m)}\mathfrak{e}_\ell(n)\sum_{k=1}^N e_{k,s}^{(m)}\mathfrak{e}_i(k) - \sum_{n=1}^N Ee_{n,s}^2\mathfrak{e}_i(n)\mathfrak{e}_\ell(n),$$

where $e_{n,s}^{(m)}$ defined in Assumption 3(b). Clearly,

$$\left|\sum_{n=1}^N Ee_{n,s}^2\mathfrak{e}_i(n)\mathfrak{e}_\ell(n)\right| \le \sup_{n\in\mathbb{N}} Ee_{n,0}^2,$$

and

$$\bar\tau_s-\bar\tau_s^{(m)} = \left\{\sum_{n=1}^N (e_{n,s}-e_{k,s}^{(m)})\mathfrak{e}_\ell(n)\right\}\sum_{k=1}^N e_{k,s}\mathfrak{e}_i(k) + \left\{\sum_{k=1}^N (e_{k,s}-e_{k,s}^{(m)})\mathfrak{e}_i(k)\right\}\sum_{n=1}^N e_{n,s}^{(m)}\mathfrak{e}_\ell(n).$$

Thus we get by the Cauchy–Schwarz inequality that

$$E|\bar\tau_0-\bar\tau_0^{(m)}|^6 \le 2^6\left[\left\{E\left|\sum_{n=1}^N (e_{n,0}-e_{k,0}^{(m)})\mathfrak{e}_\ell(n)\right|^{12} E\left|\sum_{k=1}^N e_{k,s}\mathfrak{e}_i(k)\right|^{12}\right\}^{1/2}\right.$$
$$\left. + \; E\left\{\left|\sum_{k=1}^N (e_{k,0}-e_{k,0}^{(m)})\mathfrak{e}_i(k)\right|^{12} E\left|\sum_{n=1}^N e_{n,0}^{(m)}\mathfrak{e}_\ell(n)\right|^{12}\right\}^{1/2}\right].$$

Using again Rosenthal's and Jensen's inequalities, we obtain that

$$\mathrm{E}\left|\sum_{n=1}^{N}(e_{n,0}-e_{k,0}^{(m)})\mathfrak{e}_\ell(n)\right|^{12} \le c_{4,10}\left[\sum_{n=1}^{N}\mathrm{E}|e_{n,0}-e_{k,0}^{(m)}|^{12}|\mathfrak{e}_\ell(n)|^{12} + \left\{\sum_{n=1}^{N}\mathrm{E}(e_{n,0}-e_{k,0}^{(m)})^2\mathfrak{e}_\ell^2(n)\right\}^6\right] \le c_{4,11}m^{-12\alpha},$$

and similarly

$$\mathrm{E}\left|\sum_{k=1}^{N}e_{k,0}\mathfrak{e}_i(k)\right|^{12} \le c_{4,12}\left[\sum_{k=1}^{N}\mathrm{E}|e_{k,0}|^{12}|\mathfrak{e}_i(k)|^{12} + \left\{\sum_{k=1}^{N}\mathrm{E}e_{k,0}^2\mathfrak{e}_i^2(k)\right\}^6\right] \le 2c_{4,12}\sup_{1\le k<\infty}\mathrm{E}|e_{k,0}|^{12}.$$

Thus we have

$$\mathrm{E}|\bar{\tau}_0-\bar{\tau}_0^{(m)}|^6 \le c_{4,13}m^{-6\alpha}, \tag{36}$$

and therefore Proposition 4 of [7] implies

$$\mathrm{E}\left|\sum_{s=1}^{v}\bar{\tau}_s\right|^6 \le c_{4,14}v^3. \tag{37}$$

Putting together (32)–(37) we conclude

$$\mathrm{E}\left|\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right|^6 \le c_{4,15}v^3(1+\|\boldsymbol{\gamma}\|^6+\|\boldsymbol{\gamma}\|^{12}). \tag{38}$$

Since $\{\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell : -\infty < s < \infty\}$ is a stationary sequence, (38) and the maximal inequality of [31] imply

$$\mathrm{E}\max_{v\in\{1,\ldots,z\}}\left|\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right|^6 \le c_{4,16}z^3(1+\|\boldsymbol{\gamma}\|^6+\|\boldsymbol{\gamma}\|^{12}). \tag{39}$$

Now we use (31) with $x = u(\ln T)^{1/6}$ resulting in

$$\Pr\left\{\max_{v\in\{1,\ldots,T\}}v^{-1/2}\left|\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right| > u(\ln T)^{1/6}\right\} \le c_{4,17}u^{-6},$$

implying

$$\mathrm{E}\left(\max_{v\in\{1,\ldots,T\}}v^{-1/2}\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right)^2 \le c_{4,18}(\ln T)^{1/3}.$$

This completes the proof of (30).

Next we assume that $\|\boldsymbol{\gamma}\| \to \infty$. It is easy to see that for $2 \le i \le K$

$$|Z_{N,T;i}(u)| \le \frac{1}{T}\left|\sum_{\ell\ne i,\ell\ne 1}^{N}\frac{1}{\lambda_i-\lambda_\ell}\left\{\frac{1}{(Tu)^{1/2}}\sum_{s=1}^{\lfloor Tu\rfloor}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell^\top\right\}^2\right| + \frac{1}{T}\frac{1}{\lambda_1-\lambda_2}\left\{\frac{1}{(Tu)^{1/2}}\sum_{s=2}^{\lfloor Tu\rfloor}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_1\right\}^2.$$

If $i \in \{2,\ldots,K\}$, then the proof of (38) shows that

$$\sum_{\ell\ne i,\ell\ne 1}^{N}\left\{\frac{1}{(Tu)^{1/2}}\sum_{s=1}^{\lfloor Tu\rfloor}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell\right\}^2 = O_P\{N(\ln T)^{1/3}\},$$

and therefore by Assumption 1 for any $i \in \{2,\ldots,K\}$ we have

$$\left|\sum_{\ell\ne i,\ell\ne 1}\frac{1}{\lambda_i-\lambda_\ell}\left\{\frac{1}{(Tu)^{1/2}}\sum_{s=1}^{\lfloor Tu\rfloor}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell^\top\right\}^2\right| = O_P\{N(\ln T)^{1/3}\}.$$

By (37) we have, along the lines of the proof of (31),

$$\mathrm{E}\max_{v\in\{1,\ldots,T\}}\frac{1}{v}\left\{\sum_{s=1}^{v}\mathfrak{e}_i^\top\mathbf{X}_s\mathbf{X}_s^\top\mathfrak{e}_\ell - \boldsymbol{\gamma}^\top\mathfrak{e}_i\boldsymbol{\gamma}^\top\mathfrak{e}_\ell\sum_{s=1}^{v}(\eta_s^2-1) - \right.$$
$$\left.\boldsymbol{\gamma}^\top\mathfrak{e}_i\sum_{s=1}^{v}\sum_{n=1}^{N}e_{n,s}\mathfrak{e}_\ell(n) - \boldsymbol{\gamma}^\top\mathfrak{e}_\ell\sum_{s=1}^{v}\sum_{k=1}^{N}e_{k,s}\mathfrak{e}_i(k)\right\}^2 \le c_{4,19}(\ln T)^{1/3}, \tag{40}$$

where in the last step we used (27). Also, (34) and (35) imply via the maximal inequality in [31] that

$$\mathrm{E} \max_{v\in\{1,\ldots,T\}} \left\{\frac{1}{v}\sum_{s=1}^{v}(\eta_s^2-1)\right\}^2 \le c_{4,20}(\ln T)^{1/3}, \tag{41}$$

and

$$\mathrm{E} \sup_{v\in\{1,\ldots,T\}} \frac{1}{v}\left\{\sum_{s=1}^{v}\sum_{k=1}^{N} e_{k,s}\mathfrak{e}_i(k)\right\}^2 \le c_{4,21}(\ln T)^{1/3}. \tag{42}$$

Using now (41) and (42) we conclude that

$$\frac{1}{\lambda_1-\lambda_2}\left\{\frac{1}{(Tu)^{1/2}}\sum_{s=1}^{\lfloor Tu\rfloor}\mathfrak{e}_i^\top \mathbf{X}_s\mathbf{X}_s^\top \mathfrak{e}_1\right\}^2 = \frac{(\mathfrak{e}_1^\top \boldsymbol{\gamma})^2}{\lambda_1-\lambda_2}O_P\{(\ln T)^{1/3}\}.$$

Since by Lemma 4 we have that $(\mathfrak{e}_1^\top \boldsymbol{\gamma})^2/(\lambda_1-\lambda_2)=O(1)$, the proof of (30) is complete when $i\in\{2,\ldots,K\}$. It is easy to see that, by (40) and Lemma 4,

$$\begin{aligned}\sup_{u\in[0,1]}|Z_{N,T;1}(u)| &\le \frac{1}{T}\frac{1}{\lambda_1-\lambda_2}\sup_{u\in[0,1]}\sum_{\ell=2}^{N}\left\{\frac{1}{(Tu)^{1/2}}\sum_{s=1}^{\lfloor Tu\rfloor}\mathfrak{e}_1^\top \mathbf{X}_s\mathbf{X}_s^\top \mathfrak{e}_\ell\right\}^2\\
&= \frac{1}{T}\frac{N}{\lambda_1-\lambda_2}\left[O_P(\ln T)^{1/3} + (\mathfrak{e}_1^\top \gamma)^2 \mathrm{E}\max_{v\in\{1,\ldots,T\}}\left\{v^{-1/2}\sum_{s=1}^{v}(\eta_s^2-1)\right\}^2\right.\\
&\quad\left. + \mathrm{E}\max_{2\le i\le N}\left\{v^{-1/2}\sum_{s=1}^{v}\sum_{k=1}^{N}e_{k,s}\mathfrak{e}_i(k)\right\}^2\right]\\
&= \frac{(\mathfrak{e}_1^\top \gamma)^2}{\lambda_1-\lambda_2}\frac{N(\ln T)^{1/3}}{T}\end{aligned}$$

on account of (41) and (42). According to Lemma 4 we have that $(\mathfrak{e}_1^\top \gamma)^2/(\lambda_1-\lambda_2) = O(1)$, completing the proof of Lemma 5. □

Lemma 6. *Suppose $\mathcal{H}_0$ and Assumptions 1–3 hold, and $\|\boldsymbol{\gamma}\|\to\infty$, then we have as $\min(N,T)\to\infty$, and for $c\in(0,1]$,*

$$\sup_{u\in[c,1]}|\bar{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1 - \mathfrak{e}_1^\top\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\}\mathfrak{e}_1| = O_P\left(\frac{N^2}{\|\boldsymbol{\gamma}\|^2 T}\right).$$

and

$$\sup_{u\in[0,1]}|\bar{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1 - \mathfrak{e}_1^\top\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\}\mathfrak{e}_1| = O_P\left\{\frac{N^2\ln(T)^{2/3}}{\|\boldsymbol{\gamma}\|^2 T}\right\}.$$

Proof. Let $\bar{\mathfrak{e}}_1(u)$ denote the eigenvector corresponding to the largest eigenvalue of $\mathbf{C}_{N,T}(u)$, multiplied by either 1 or −1 so that $\mathrm{sign}\{\langle\bar{\mathfrak{e}}_1(u),\mathfrak{e}_1\rangle\}>0$. Then according to the definitions of $\bar{\lambda}_1(u)$, $\bar{\mathfrak{e}}_1(u)$, λ_1, and $\mathfrak{e}_1$,

$$\begin{aligned}&[(\lfloor Tu\rfloor/T)\mathbf{C}+\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\}]\times[\mathfrak{e}_1+\{\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\}]\\
&\quad=[(\lfloor Tu\rfloor/T)\lambda_1+\{\bar{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1\}]\times[\mathfrak{e}_1+\{\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\}].\end{aligned}$$

By rearranging these terms we have that

$$\begin{aligned}\{\bar{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1\}\mathfrak{e}_1 &= \{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\}\mathfrak{e}_1+(\lfloor Tu\rfloor/T)\mathbf{C}\{\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\}\\
&\quad-(\lfloor Tu\rfloor/T)\lambda_1\{\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\}+G_{N,T}(u),\end{aligned} \tag{43}$$

where

$$G_{N,T}(u)=\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\}\{\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\}-\{\bar{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1\}\times\{\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\}.$$

We then have according to the triangle and Cauchy–Schwarz inequalities that

$$\begin{aligned}|\mathfrak{e}_1^\top G_{N,T}(u)| &\le |\mathfrak{e}_1^\top\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\}\times\{\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\}|+|\{\bar{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1\}\times\mathfrak{e}_1^\top\{\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\}|\\
&\le \|\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\|\times\|\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\|+|\bar{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1|\times\|\bar{\mathfrak{e}}_1(u)-\mathfrak{e}_1\|.\end{aligned} \tag{44}$$

We have, using the results in Chapter 7 of [11], that

$$|\bar{\lambda}_1(u) - (\lfloor Tu \rfloor/T)\lambda_1| \le \|\mathbf{C}_{N,T}(u) - (\lfloor Tu \rfloor/T)\mathbf{C}\|,$$

from which it follows along with (44) that

$$\sup_{u\in[c,1]} |\mathfrak{e}_1^\top G_{N,T}(u)| \le c_{5,1} \sup_{u\in[c,1]} \|\mathbf{C}_{N,T}(u) - (\lfloor Tu \rfloor/T)\mathbf{C}\| \sup_{u\in[c,1]} \|\bar{\mathfrak{e}}_1(u) - \mathfrak{e}_1\|.$$

Following the proof of Lemma 1, one can verify that

$$\sup_{u\in[c,1]} \|\mathbf{C}_{N,T}(u) - (\lfloor Tu \rfloor/T)\mathbf{C}\| = O(NT^{-1/2}). \tag{45}$$

Furthermore according to the inequality of [11],

$$\sup_{u\in[c,1]} \|\bar{\mathfrak{e}}_1(u) - \mathfrak{e}_1\| \le c_{5,2} \frac{1}{\lambda_1 - \lambda_2} \sup_{u\in[c,1]} \|\mathbf{C}_{N,T}(u) - (\lfloor Tu \rfloor/T)\mathbf{C}\| = O\left(\frac{N}{\|\boldsymbol{\gamma}\|^2 T^{1/2}}\right),$$

since, by Lemma 4, $1/(\lambda_1 - \lambda_2) = O(\|\boldsymbol{\gamma}\|^{-2})$. It follows that

$$\sup_{u\in[c,1]} |\mathfrak{e}_1^\top G_{N,T}(u)| = O_P\left(\frac{N^2}{\|\boldsymbol{\gamma}\|^2 T^1}\right). \tag{46}$$

By multiplying the left- and right-hand sides of (43) by $\mathfrak{e}_1^\top$, we obtain

$$\bar{\lambda}_1(u) - (\lfloor Tu \rfloor/T)\lambda_1 = \mathfrak{e}_1^\top \{\mathbf{C}_{N,T}(u) - (\lfloor Tu \rfloor/T)\mathbf{C}\}\mathfrak{e}_1 + \mathfrak{e}_1^\top G_{N,T}(u),$$

from which the first part of the lemma now follows from (46). The second part follows by replacing the supremum to be taken over the whole interval, and replacing (45) with the bound achieved in (41). □

Using the definition of $\mathbf{C}_{N,T}(u)$ and (1) we get for any $i \in \{1, \ldots, K\}$,

$$\begin{aligned} T\mathfrak{e}_i^\top(\mathbf{C}_{N,T}(u) - (\lfloor Tu \rfloor/T)\mathbf{C})\mathfrak{e}_i &= (\mathfrak{e}_i^\top \boldsymbol{\gamma})^2 \sum_{t=1}^{\lfloor Tu \rfloor}(\eta_t^2 - 1) + 2\mathfrak{e}_i^\top \boldsymbol{\gamma} \sum_{t=1}^{\lfloor Tu \rfloor} \eta_t \sum_{\ell=1}^{N} \mathfrak{e}_i(\ell) e_{\ell,t} \\ &\quad + \sum_{t=1}^{\lfloor Tu \rfloor} \left\{\sum_{\ell=1}^{N} \mathfrak{e}_i(\ell) e_{\ell,t}\right\}^2 - \lfloor Tu \rfloor \sum_{\ell=1}^{N} \mathfrak{e}_i^2(\ell)\sigma_\ell^2. \end{aligned}$$

For each $i \in \{1, \ldots, K\}$

$$D_{N,T}(u) = \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor Tu \rfloor}(\eta_t^2 - 1), \quad F_{N,T;i}(u) = \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor Tu \rfloor} \eta_t \sum_{\ell=1}^{N} \mathfrak{e}_i(\ell) e_{\ell,t},$$

and

$$G_{N,T;i}(u) = \frac{1}{T^{1/2}} \left[\sum_{t=1}^{\lfloor Tu \rfloor} \left\{\sum_{\ell=1}^{N} \mathfrak{e}_i(\ell) e_{\ell,t}\right\}^2 - \lfloor Tu \rfloor \sum_{\ell=1}^{N} \mathfrak{e}_i^2(\ell)\sigma_\ell^2\right],$$

Lemma 7. *If (1) and Assumptions 1–3 hold, then* $\{D_{N,T}(u), F_{N,T;i}(u), G_{N,T;i}(u) : 0 \le u \le 1, 1 \le i \le K\}$ *converges in* $\mathcal{D}^{2K+1}[0, 1]$ *to the Gaussian process* $\boldsymbol{\Gamma}(u) = (\Gamma_1(u), \ldots, \Gamma_{2K+1}(u))^\top$ *defined for all* $u \in [0, 1]$, *where* $\mathrm{E}\boldsymbol{\Gamma}(u) = \mathbf{0}$, *and*

$$\mathrm{E}\boldsymbol{\Gamma}(u)\boldsymbol{\Gamma}^\top(u') = \min(u, u') \begin{pmatrix} \sigma_\eta^2 & \mathbf{0}^\top & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{V}_2 & \mathbf{O} \\ \mathbf{0} & \mathbf{O} & \mathbf{V}_3 \end{pmatrix}$$

Proof. First, for each $i \in \{1, \ldots, K\}$, we define the m-dependent processes

$$D_{N,T}^{(m)}(u) = \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor Tu \rfloor}\{(\eta_t^{(m)})^2 - 1\}, \quad F_{N,T;i}^{(m)}(u) = \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor Tu \rfloor} \eta_t \sum_{\ell=1}^{N} \mathfrak{e}_i(\ell) e_{\ell,t}^{(m)},$$

and

$$G_{N,T;i}^{(m)}(u) = \frac{1}{T^{1/2}} \left[\sum_{t=1}^{\lfloor Tu \rfloor} \left\{\sum_{\ell=1}^{N} \mathfrak{e}_i(\ell) e_{\ell,t}^{(m)}\right\}^2 - \lfloor Tu \rfloor \sum_{\ell=1}^{N} \mathfrak{e}_i^2(\ell)\sigma_\ell^2\right],$$

where $\eta_t^{(m)}$ and $e_{\ell,t}^{(m)}$ are defined in Assumption 3(a) and Assumption 3(b), respectively. We show that for any $x > 0$

$$\lim_{m\to\infty}\limsup_{T\to\infty}\Pr\{|D_{N,T}(u) - D_{N,T}^{(m)}(u)| > x\} = 0, \tag{47}$$

$$\lim_{m\to\infty}\limsup_{T\to\infty}\Pr\{|F_{N,T;i}(u) - F_{N,T;i}^{(m)}(u)| > x\} = 0, \tag{48}$$

and

$$\lim_{m\to\infty}\limsup_{T\to\infty}\Pr\{|G_{N,T;i}(u) - G_{N,T;i}^{(m)}(u)| > x\} = 0, \tag{49}$$

for all $u \in (0, 1]$ and $i \in \{1, \ldots, K\}$. It follows from Assumption 3(a) and the Cauchy–Schwarz inequality that

$$\mathrm{E}|\eta_0^2 - (\eta_0^{(m)})^2|^6 = \mathrm{E}\{|\eta_0 + \eta_0^{(m)}|\,|\eta_0 - \eta_0^{(m)}|\}^6 \le 2^4(\mathrm{E}\eta_0^{12})^{1/2}(\mathrm{E}|\eta_0 - \eta^{(m)}|^{12})^{1/2} \le c_{6,1}m^{-6\alpha}. \tag{50}$$

By stationarity, we get that

$$\begin{aligned}\mathrm{var}\left[T^{-1/2}\sum_{s=1}^{\lfloor Tu\rfloor}\{\eta_s^2 - (\eta_s^{(m)})^2\}^2\right] &\le \frac{1}{T}\sum_{s=1}^{T}\mathrm{E}[\eta_s^2 - \{\eta_s^{(m)}\}^2] + 2\sum_{s=1}^{T}\mathrm{E}[\eta_0^2 - \{\eta_0^{(m)}\}^2]\times[\eta_s^2 - \{\eta_s^{(m)}\}^2]\\ &\le \mathrm{E}[\eta_0^2 - \{\eta_0^{(m)}\}^2]^2 + 2\sum_{s=1}^{T}|\mathrm{E}[\eta_0^2 - \{\eta_0^{(m)}\}^2]\times[\eta_s^2 - \{\eta_s^{(m)}\}^2]|.\end{aligned}$$

Since $\eta_0^2 - (\eta_0^{(m)})^2$ is independent of $\eta_s^{(m)}$, if $s > m$, we obtain that

$$\sum_{s=m+1}^{T}|\mathrm{E}[\eta_0^2 - \{\eta_0^{(m)}\}^2]\times[\eta_s^2 - \{\eta_s^{(m)}\}^2]| \le \sum_{s=m+1}^{T}|\mathrm{E}(\eta_0^2 - 1)\eta_s^2| + \sum_{s=m+1}^{T}|\mathrm{E}[\{\eta_0^{(m)}\}^2 - 1]\eta_s^2|.$$

The independence of η_0 and $\eta_s^{(s)}$, (50), and Hölder's inequality yield

$$\sum_{s=m+1}^{T}|\mathrm{E}(\eta_0^2 - 1)\eta_s^2| = \sum_{s=m+1}^{T}|\mathrm{E}(\eta_0^2 - 1)[\eta_s^2 - \{\eta_s^{(s)}\}^2]| \le \sum_{s=m+1}^{\infty}(\mathrm{E}|\eta_0^2 - 1|^{6/5})^{5/6}[\mathrm{E}[\eta_0^2 - \{\eta_0^{(s)}\}^2]^6]^{1/6} \le c_{6,2}m^{-(\alpha-1)}$$

with $c_{6,2} = \{c_{6,1}/(\alpha - 1)\}(\mathrm{E}|\eta_0^2 - 1|^{6/5})^{5/6}$. The same argument yields

$$\sum_{s=m+1}^{T}|[\mathrm{E}[\{\eta_0^{(m)}\}^2] - 1]\eta_s^2| \le c_{6,2}m^{-(\alpha-1)}.$$

On the other hand, applying again (50) and the Cauchy–Schwarz inequality we conclude

$$\sum_{s=1}^{m}|\mathrm{E}[\eta_0^2 - \{\eta_0^{(m)}\}^2]\times[\eta_s^2 - \{\eta_s^{(m)}\}^2]| \le \sum_{s=1}^{m}\mathrm{E}[\eta_0^2 - \{\eta_0^{(m)}\}^2]^2 \le c_{6,1}m^{-(\alpha-1)}.$$

Chebyshev's inequality now implies (47). The proofs of (48) and (49) go along the lines of (47), we only need to replace (50) with (33) and (36), respectively. Next we show that for each m, $\{D_{N,T}^{(m)}(u), F_{N,T;i}^{(m)}(u), G_{N,T;i}^{(m)}(u) : 0 \le u \le 1, 1 \le i \le K\}$ converges in $\mathcal{D}^{2K+1}[0, 1]$ to the Gaussian process $\boldsymbol{\Gamma}^{(m)}(u) = (\Gamma_1^{(m)}(u), \ldots, \Gamma_{2K+1}^{(m)}(u))^\top$ defined for all $u \in [0, 1]$, with $\mathrm{E}\boldsymbol{\Gamma}^{(m)}(u) = \mathbf{0}$, and

$$\mathrm{E}\boldsymbol{\Gamma}^{(m)}(u)(\boldsymbol{\Gamma}^{(m)})^\top(u') = \min(u, u')\begin{pmatrix}(\sigma_\eta^2)^{(m)} & \mathbf{0}^\top & \mathbf{0}^\top\\ \mathbf{0} & \mathbf{V}_2^{(m)} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{V}_3^{(m)}\end{pmatrix}$$

with

$$(\sigma_\eta^2)^{(m)} = \sum_{\ell=-m}^{m}\mathrm{cov}\{(\eta_0^{(m)})^2, (\eta_\ell^{(m)})^2\}, \tag{51}$$

$$\mathbf{V}_2^{(m)} = \left\{\sum_{s=-m}^{m}\lim_{N\to\infty}\sum_{k=1}^{N}\mathfrak{e}_i(k)\mathfrak{e}_j(k)\mathrm{cov}\{\eta_0^{(m)}, \eta_s^{(m)}\}\mathrm{cov}\{e_{k,0}^{(m)}, e_{k,s}^{(m)}\} : 1 \le i, j \le K\right\}, \tag{52}$$

and

$$\mathbf{V}_3^{(m)} = \left\{ \sum_{s=-m}^{m} \lim_{N\to\infty} \left[\sum_{k=1}^{N} \mathfrak{e}_i^2(k)\mathfrak{e}_j^2(k)\mathrm{cov}\Big[\{e_{k,0}^{(m)}\}^2, \{e_{k,s}^{(m)}\}^2\Big] + 2\left[\sum_{k=1}^{N} \mathfrak{e}_i(k)\mathfrak{e}_j(k)\mathrm{cov}\{e_{k,0}^{(m)}, e_{k,s}^{(m)}\}\right]^2 \right.\right.$$
$$\left.\left. -2\sum_{k=1}^{N} \mathfrak{e}_i^2(k)\mathfrak{e}_j^2(k)\Big[\mathrm{cov}\Big\{e_{k,0}^{(m)}, e_{k,s}^{(m)}\Big\}\Big]^2 \right] : 1 \le i,j \le K \right\}. \tag{53}$$

Let $0 \le u_1 < \cdots < u_M \le 1$ and $\mu_{i,k,\ell}$ for all $i \in \{1,\ldots,M\}$ and $k \in \{1,\ldots,K\}$. We can write

$$\sum_{k=1}^{M} \mu_{k,1,1}\{D_{N,T}^{(m)}(u_k) - D_{N,T}^{(m)}(u_{k-1})\} + \sum_{k=1}^{M}\sum_{i=1}^{K} \mu_{k,2,i}\{F_{N,T,i}^{(m)}(u_k) - F_{N,T,i}^{(m)}(u_{k-1})\}$$
$$+ \sum_{k=1}^{M}\sum_{i=1}^{K} \mu_{k,3,i}\{G_{N,T,i}^{(m)}(u_k) - G_{N,T,i}^{(m)}(u_{k-1})\} = \mathcal{S}_1 + \cdots + \mathcal{S}_M,$$

where, for each $i \in \{1,\ldots,M\}$,

$$\mathcal{S}_k = \sum_{s=\lfloor Tu_{i-1}\rfloor+1}^{\lfloor Tu_i \rfloor} \xi_{N,T;s}(k).$$

The variables $\xi_{N,T;s}(k)$, $\lfloor Tu_{k-1}\rfloor + 1 \le s \le \lfloor Tu_k \rfloor$, $1 \le k \le M$ are m-dependent and therefore $T^{-1/2}\mathcal{S}_1, T^{-1/2}\mathcal{S}_2, \ldots, T^{-1/2}\mathcal{S}_M$ are asymptotically independent. Hence we need only show the asymptotic normality of $T^{-1/2}\mathcal{S}_k$ for all $k \in \{1,\ldots,M\}$. For every fixed k the variables $\xi_{N,T;s}(k)$, $\lfloor Tu_{k-1}\rfloor + 1 \le s \le \lfloor Tu_k\rfloor$ form an m-dependent stationary sequence with zero mean,

$$\lim_{T\to\infty} \mathrm{var}(T^{-1/2}\mathcal{S}_k) = \mathrm{var}\left[\mu_{k,1,1}\Gamma_1^{(m)}(u_k) - \{\Gamma_1^{(m)}(u_{k-1})\} + \sum_{i=1}^{K} \mu_{k,2,i}\{\Gamma_{i+1}^{(m)}(u_k) - \Gamma_{i+1}^{(m)}(u_{k-1})\} \right.$$
$$\left. + \sum_{i=1}^{K} \mu_{k,3,i}\{\Gamma_{i+K+1}^{(m)}(u_k) - \Gamma_{i+K+1}^{(m)}(u_{k-1})\} \right]$$

and $\mathrm{E}|\xi_{N,T;s}(k)|^3 \le C_1$, where $C_{1,1}$ does not depend on N nor on T. Due to the m-dependence, these properties imply the asymptotic normality of $T^{-1/2}\mathcal{S}_k$. Applying the Cramér–Wold device [8], we get that the finite-dimensional distributions of $\{D_{N,T}^{(m)}(u), F_{N,T;i}^{(m)}(u), G_{N,T;i}^{(m)}(u) : 0 \le u \le 1, 1 \le i \le K\}$ converge to that of $\mathbf{\Gamma}^{(m)}(u)$. Since $\|\mathbf{V}^{(m)} - \mathbf{V}\| \to 0$ as $T \to \infty$, and $\mathbf{\Gamma}(u)$ and $\mathbf{\Gamma}^{(m)}(u)$ are Gaussian processes we conclude that $\mathbf{\Gamma}^{(m)}(u)$ converges in $\mathcal{D}^{2K+1}[0,1]$ to $\mathbf{\Gamma}(u)$. On account of (47)–(49) we obtain that the finite-dimensional distributions of $\{D_{N,T}(u), F_{N,T;i}(u), G_{N,T;i}(u) : 0 \le u \le 1, 1 \le i \le K\}$ converge to that of $\mathbf{\Gamma}(u)$. It is shown in the proof of Lemma 1 that

$$\mathrm{E}\left|\sum_{t=1}^{v}(\eta_t^2 - 1)\right|^3 \le c_{6,3}v^{3/2}, \quad \mathrm{E}\left|\sum_{t=1}^{v}\eta_t\sum_{\ell=1}^{N}\mathfrak{e}_i(\ell)e_{\ell,t}\right|^3 \le c_{6,4}v^{3/2}$$

and

$$\mathrm{E}\left|\sum_{t=1}^{v}\left\{\sum_{\ell=1}^{N}\mathfrak{e}_i(\ell)e_{\ell,t}\right\}^2 - v\sum_{\ell=1}^{N}\mathfrak{e}_i^2(\ell)\sigma_\ell^2\right|^3 \le c_{6,5}v^{3/2}.$$

Due to the stationarity of $\eta, e_{i,t},\ i \in \{1,\ldots,N\}$, the tightness follows from Theorem 8.4 of [8]. □

Proof of Theorem 2. First we consider the case when $\|\boldsymbol{\gamma}\| \to \infty$. Lemmas 2 and 3 yield

$$\sup_{u\in[0,1]} |T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\{\tilde{\lambda}_1(u) - (\lfloor Tu\rfloor/T)\lambda_1\} - T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\mathfrak{e}_1^\top\Big\{\mathbf{C}_{N,T}(u) - (\lfloor Tu\rfloor/T)\mathbf{C}\Big\}\mathfrak{e}_1| = O_P\left\{\left(\frac{N^2}{T} + \frac{N}{T^{1/2}}\right)\frac{(\ln T)^{1/3}}{\|\boldsymbol{\gamma}\|^2}\right\}.$$

In addition, we have from Lemma 6 that

$$\sup_{u\in[0,1]} |T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\{\tilde{\lambda}_1(u) - (\lfloor Tu\rfloor/T)\lambda_1\} - T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\mathfrak{e}_1^\top\{\mathbf{C}_{N,T}(u) - (\lfloor Tu\rfloor/T)\mathbf{C}\}\mathfrak{e}_1| = O_P\left\{\frac{N^2(\ln T)^{1/3}}{\|\boldsymbol{\gamma}\|^4 T^{1/2}}\right\}.$$

Therefore, under (5),

$$\sup_{u\in[0,1]} |T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\{\tilde{\lambda}_1(u) - (\lfloor Tu\rfloor/T)\lambda_1\} - T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\mathfrak{e}_1^\top\{\mathbf{C}_{N,T}(u) - (\lfloor Tu\rfloor/T)\mathbf{C}\}\mathfrak{e}_1| = o_P(1).$$

Thus Lemma 7 implies that

$$\sup_{u\in[0,1]}\left|T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\{\tilde{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1\}-\frac{(\mathfrak{e}_1^\top\boldsymbol{\gamma})^2}{\|\boldsymbol{\gamma}\|^2}D_{N,T}(u)\right|=o_P(1).$$

According to Lemma 7, $\sup_{u\in[0,1]}|D_{N,T}(u)|=O_P(1)$ and since $(\mathfrak{e}_1^\top\boldsymbol{\gamma})^2/\|\boldsymbol{\gamma}\|^2\to 1$ by Lemma 4, we conclude

$$\sup_{u\in[0,1]}|T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\{\bar{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1\}-D_{N,T}(u)|=o_P(1).$$

Therefore by Lemma 7 we obtain that $T^{1/2}|\,\|\boldsymbol{\gamma}\|^{-2}\{\tilde{\lambda}_1(u)-u\lambda_1\}$ converges weakly in $\mathcal{D}[0,1]$ to $\sigma_\eta W(u)$, which establishes the first part of the theorem.

When $\|\boldsymbol{\gamma}\|=O(1)$, we again have by Lemmas 2, 3, and 5 that

$$\sup_{u\in[0,1]}|T^{1/2}\{\tilde{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1\}-T^{1/2}\mathfrak{e}_1^\top\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\}\mathfrak{e}_1|=O_P\{N(\ln T)^{1/3}/T^{1/2}\}=o_P(1),$$

under (6). Also,

$$\begin{aligned}\sup_{u\in[0,1]}|T^{1/2}\mathfrak{e}_1^\top\left\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\right\}\mathfrak{e}_1-G_{N,T;1}(u)|&\le(\mathfrak{e}_1^\top\boldsymbol{\gamma})^2\sup_{u\in[0,1]}|D_{N,T}(u)|+2|\mathfrak{e}_1^\top\boldsymbol{\gamma}|\sup_{u\in[0,1]}|F_{N,T;1}(u)|\\&=O_P(1)\{(\mathfrak{e}_1^\top\boldsymbol{\gamma})^2+|\mathfrak{e}_1^\top\boldsymbol{\gamma}|\},\end{aligned}$$

since by Lemma 7

$$\sup_{u\in[0,1]}|D_{N,T}(u)|=O_P(1)\quad\text{and}\quad\sup_{u\in[0,1]}|F_{N,T;1}(u)|=O_P(1).$$

By the Cauchy–Schwarz inequality we have that $|\mathfrak{e}_i^\top\boldsymbol{\gamma}|\le\|\boldsymbol{\gamma}\|=O(1)$ and therefore

$$\sup_{u\in[0,1]}|T^{1/2}\mathfrak{e}_1^\top\left\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\right\}\mathfrak{e}_1-G_{N,T;1}(u)|=o_P(1).$$

The weak convergence of the process $G_{N,T;1}(u)$, is proven in Lemma 7, which completes the proof of the second part of Theorem 2. □

Proof of Theorem 1. This follows precisely as Theorem 2 by only replacing Lemma 5 with the result that for all $c>0$

$$\max_{i\in\{1,\ldots,K\}}\sup_{u\in[c,1]}|Z_{N,T;i}(u)|=O_P\left(N/T\right),$$

which follows from (39) and Markov's inequality. □

Proof of Theorem 8. By Lemmas 2–5 we have that

$$\sup_{u\in[0,1]}|T^{1/2}\{\tilde{\lambda}_i(u)-(\lfloor Tu\rfloor/T)\lambda_i\}-T^{1/2}\mathfrak{e}_i^\top\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\}\mathfrak{e}_i|=o_P(1).$$

Also,

$$\begin{aligned}&\sup_{u\in[0,1]}\left|T^{1/2}\mathfrak{e}_i^\top\left(\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\right)\mathfrak{e}_i-G_{N,T;i}(u)\right|\\&\quad\le(\mathfrak{e}_i^\top\boldsymbol{\gamma})^2\sup_{u\in[0,1]}|D_{N,T}(u)|+2|\mathfrak{e}_i^\top\boldsymbol{\gamma}|\sup_{u\in[0,1]}|F_{N,T;i}(u)|=O_P(1)((\mathfrak{e}_i^\top\boldsymbol{\gamma})^2+|\mathfrak{e}_i^\top\boldsymbol{\gamma}|),\end{aligned}$$

since by Lemma 7

$$\sup_{u\in[0,1]}|D_{N,T}(u)|=O_P(1)\quad\text{and}\quad\sup_{u\in[0,1]}|F_{N,T;i}(u)|=O_P(1).$$

By the Cauchy–Schwarz inequality we have that $|\mathfrak{e}_i^\top\boldsymbol{\gamma}|\le\|\boldsymbol{\gamma}\|$ and therefore

$$\sup_{u\in[0,1]}|T^{1/2}\mathfrak{e}_i^\top\left\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\right\}\mathfrak{e}_i-G_{N,T;i}(u)|=o_P(1).$$

The weak convergence of $G_{N,T;i}(u)$, defined for all $u\in[0,1]$ for all $i\in\{1,\ldots,K\}$, is proven in Lemma 7, which completes the proof of Theorem 2. □

Proof of Theorem 9. Lemmas 2 and 3 yield

$$\sup_{u\in[0,1]}|T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\{\tilde{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1\}-T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\mathfrak{e}_1^\top\{\mathbf{C}_{N,T}(u)-(\lfloor Tu\rfloor/T)\mathbf{C}\}\mathfrak{e}_1|=o_P(1).$$

Thus Lemma 7 yields

$$\sup_{u\in[0,1]}\left|T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\{\tilde{\lambda}_1(u)-(\lfloor Tu\rfloor/T)\lambda_1\}-\frac{(\mathfrak{e}_1^\top\boldsymbol{\gamma})^2}{\|\boldsymbol{\gamma}\|^2}D_{N,T}(u)\right|=o_P(1).$$

According to Lemma 7, $\sup_{u\in[0,1]}|D_{N,T}(u)| = O_P(1)$ and since $(\mathfrak{e}_1^\top \boldsymbol{\gamma})^2/\|\boldsymbol{\gamma}\|^2 \to 1$ by Lemma 4, we conclude again that

$$\sup_{u\in[0,1]} |T^{1/2}\|\boldsymbol{\gamma}\|^{-2}\{\tilde{\lambda}_1(u) - (\lfloor Tu\rfloor/T)\lambda_1\} - D_{N,T}(u)| = o_P(1). \tag{54}$$

Lemmas 4 and 5 imply

$$\sup_{u\in[0,1]} |T^{1/2}\{\tilde{\lambda}_i(u) - u\lambda_i\} - \{(\mathfrak{e}_i^\top \boldsymbol{\gamma})^2 D_{N,T}(u) + 2\mathfrak{e}_i^\top \boldsymbol{\gamma} F_{N,T;i}(u) + G_{N,T;i}(u)\}| = o_P(1). \tag{55}$$

Combining (54) and (55) with Lemma 7, we obtain that $\{T^{1/2}|\,\|\boldsymbol{\gamma}\|^{-2}\{\tilde{\lambda}_1(u) - u\lambda_1\}, T^{1/2}(\tilde{\lambda}_i(u) - u\lambda_i) : 2 \le i \le K\}$ converges weakly in $\mathcal{D}^K[0, 1]$ to $\boldsymbol{\Gamma}^0(u) = (\Gamma_1^0(u), \ldots, \Gamma_K^0(u))^\top$, where $\Gamma_1^0(u) = \Gamma_1(u)$ and $\Gamma_i^0(u) = a_i^2\Gamma_1(u) + 2a_i\Gamma_{i+1}(u) + \Gamma_{i+K+1}(u)$ for all $i \in \{2, \ldots, K\}$. The computation of the covariance function of $\boldsymbol{\Gamma}^0(u)$ finishes the proof of Theorem 9. □

Acknowledgments

We would like to thank the Associate Editor, and two referees whose careful reading and insightful comments led to a much improved paper. We would also like to express our gratitude to the Editor-in-Chief for collaborating with us to streamline the presentation of this work. The second author was partially supported by a Discovery grant and an Accelerator grant from the Natural Sciences and Engineering Research Council of Canada. We would also like to thank Yijun Xie for his help in producing Fig. 3 and Table 1.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2018.07.001.

References

[1] D. Andrews, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59 (1991) 817–858.
[2] A. Aue, S. Hörmann, L. Horváth, M. Reimherr, Break detection in the covariance structure of multivariate time series models, Ann. Statist. 37 (2009) 4046–4087.
[3] A. Aue, D. Paul, Random matrix theory in statistics: A review, J. Statist. Plann. Inference 150 (2014) 1–29.
[4] J. Bai, Inferential theory for factor models of large dimensions, Econometrica 71 (2003) 135–171.
[5] J. Bai, Common breaks in means and variances for panel data, J. Econom. 157 (2010) 78–92.
[6] J. Bai, S. Ng, Determining the number of factors in approximate factor models, Econometrica 70 (2002) 191–221.
[7] I. Berkes, S. Hörmann, J. Schauer, Split invariance principles for stationary processes, Ann. Probab. 39 (2011) 2441–2473.
[8] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[9] J. Breitung, S. Eickmeier, Testing for structural breaks in dynamic factor models, J. Econom. 163 (2011) 71–84.
[10] L. Chen, J.J. Dolado, J. Gonzalo, Detecting big structural breaks in large factor models, J. Econom. 180 (2014) 30–48.
[11] N. Dunford, J.T. Schwartz, Linear Operators, General Theory (Part 1), Wiley, New York, 1988.
[12] J. Fan, Y. Fan, J. Lv, High dimensional covariance matrix estimation using a factor model, J. Econom. 147 (2008) 186–197.
[13] P. Galeano, D. Peña, Covariance changes detection in multivariate time series, J. Statist. Plann. Inference 137 (2007) 194–221.
[14] R. Gürkaynak, B. Sack, J. Wright, The US treasury yield curve: 1961 to the present, J. Monetary Econom. 54 (2007) 2291–2304.
[15] P. Hall, M. Hosseini-Nasab, Theory for high-order bounds in functional principal components analysis, Math. Proc. Cambridge Philos. Soc. 146 (2009) 225–256.
[16] A. Halungaa, D. Osborn, Ratio-based estimators for a change point in persistence, J. Econom. 171 (2012) 24–31.
[17] X. Han, A. Inoue, Tests for parameter instability in dynamic factor models, Econometric Theory 31 (2015) 1117–1152.
[18] L. Horváth, M. Hušková, Change-point detection in panel data, J. Time Series Anal. 33 (2012) 631–648.
[19] L. Horváth, G. Rice, Online Supplement for Asymptotics for empirical eigenvalue processes in high-dimensional linear factor models (2018).
[20] I. Johnstone, Multivariate analysis and jacobi ensembles: largest eigenvalue, Tracy–Widom limits and rates of convergence, Ann. Statist. 36 (2008) 2638–2716.
[21] S. Jung, J.S. Marron, PCA consistency in high dimension low sample size context, Ann. Statist. 37 (2009) 4104–4130.
[22] C. Kao, L. Trapani, G. Urga, Testing for Breaks in Cointegrated Panels, Technical Report, Center for Policy Research, 157 (2012), https://surface.syr.edu/cpr/157.
[23] C. Kao, L. Trapani, G. Urga, Testing for instability in covariance structures, Bernoulli 24 (2018) 740–771.
[24] M. Kejriwal, Test of a mean shift with good size and monotonic power, Econom. Lett. 102 (2015) 78–82.
[25] G. Keogh, S. Sharifi, H. Ruskin, M. Crane, Epochs in market sector index data–empirical or optimistic? in: The Application of Econophysics, Springer, New York, 2004, pp. 83–89.
[26] D. Kim, Estimating a common deterministic time trend break in large panels with cross sectional dependence, J. Econom. 164 (2011) 310–330.
[27] D. Kim, Common breaks in time trends for large panel data with a factor structure, Econom. J. 17 (2014) 301–337.
[28] D. Li, J. Qian, L. Su, Panel data models with interactive fixed effects and multiple structural breaks, J. Amer. Statist. Assoc. 111 (2016) 1804–1819.
[29] H. Markowitz, Portfolio selection, J. Finance 7 (1952) 77–91.
[30] H. Markowitz, The optimization of a quadratic function subject to linear constraints, Naval Res. Logist. 3 (1956) 111–133.
[31] F. Móricz, R. Serfling, W. Stout, Moment and probability bounds with quasi-superadditive structure for the maximal partial sum, Ann. Probab. 10 (1982) 1032–1040.
[32] A. Onatski, Asymptotics of the principal components estimator of large factor models with weakly influential factors, J. Econom. 168 (2012) 244–258.
[33] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, Statist. Sinica 17 (2007) 1617–1642.
[34] P. Perron, A test for changes in a polynomial trend function for a dynamic time series, in: Research MemorandUm No. 363, Econometric Research Program, Princeton University, Princeton, NJ, 1991.
[35] V.V. Petrov, Limit Theorems of Probability Theory, Clarendon Press, Oxford, 1995.

[36] J. Qian, L. Su, Shrinkage Estimation of Common Breaks in Panel Data Models Via Adaptive Group Fused Lasso, Working Paper, Singapore Management University, 2014.
[37] R Development Core Team (2008). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.
[38] X. Shao, Self-normalization for time series: A review of recent developments, J. Amer. Statist. Assoc. 110 (2015) 1797–1817.
[39] A. Taniguchi, Y. Kakizawa, Asymptotic Theory of Statistical Inference for Time Series, Springer, New York, 2000.
[40] T.J. Vogelsang, Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series, J. Econom. 88 (1999) 283–299.
[41] W. Wang, J. Fan, Asymptotics of empirical eigenstructure for high dimensional spiked covariance, Ann. Statist. 45 (2017) 1342–1374.
[42] D. Wied, W. Krämer, H. Dehling, Testing for a change in correlation at an unknown point in time using an extended functional delta method, Econometric Theory 28 (2012) 570–589.
[43] W. Wu, Nonlinear system theory: another look at Dependence, Proc. Natl. Acad. Sci. USA 102 (2005) 14150–14154.
[44] Y. Yamamoto, S. Tanaka, Testing for factor loading structural change under common breaks, J. Econom. 189 (2015) 187–206.
[45] A. Zeileis, Object-oriented computation of sandwich estimators, J. Stat. Softw. 16 (2006) 1–16.
[46] Z. Zhou, Heteroscedastic and autocorrelation robust structural change detection, J. Amer. Statist. Assoc. 108 (2013) 726–740.
[47] I.I. Zovko, J.D. Farmer, Correlations and clustering in the trading of members of the london stock exchange, in: Complexity, Metastability and Nonextensivity: An International Conference AIP Conference Proceedings, Springer, New York, 2007.

Further Reading

[1] A. Ang, J. Chen, Asymmetric correlation of equity portfolios, J. Financ. Econom. 63 (2002) 443–494.
[2] A. Aue, L. Horváth, Structural breaks in time series, J. Time Series Anal. 34 (2013) 1–16.
[3] G. Chamberlain, M. Rothschild, Funds factors, and diversification in arbitrage pricing models, Econometrica 51 (1983) 1305–1324.
[4] A. Rényi, On the theory of order statistics, Acta Math. Acad. Sci. Hungar. 4 (1953) 191–227.
[5] G.R. Shorack, J.A. Wellner, Empirical Processes with Applications to Statistics, Wiley, New York, 1986.
[6] L. Trapani, A randomised sequential procedure to determine the number of factors, J. Amer. Statist. Assoc. (2018) (in press).