# The fluctuation of eigenvalues in factor models

Fanglin Bao, and Bo Zhang ✉

*Department of Statistics and Finance, International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230026, China*

✉Correspondence: Bo Zhang, E-mail: wbchpmp@ustc.edu.cn

## Graphical abstract



*The randomization algorithm is used to construct the test statistics of the model structure.*

## Public summary

■ Study the fluctuation characteristics of eigenvalues in factor models.

■ Construct the test statistic using a randomization algorithm.

■ Check for breaks in the structure of factor models, involving changes in factor loadings and increases in the number of factors.
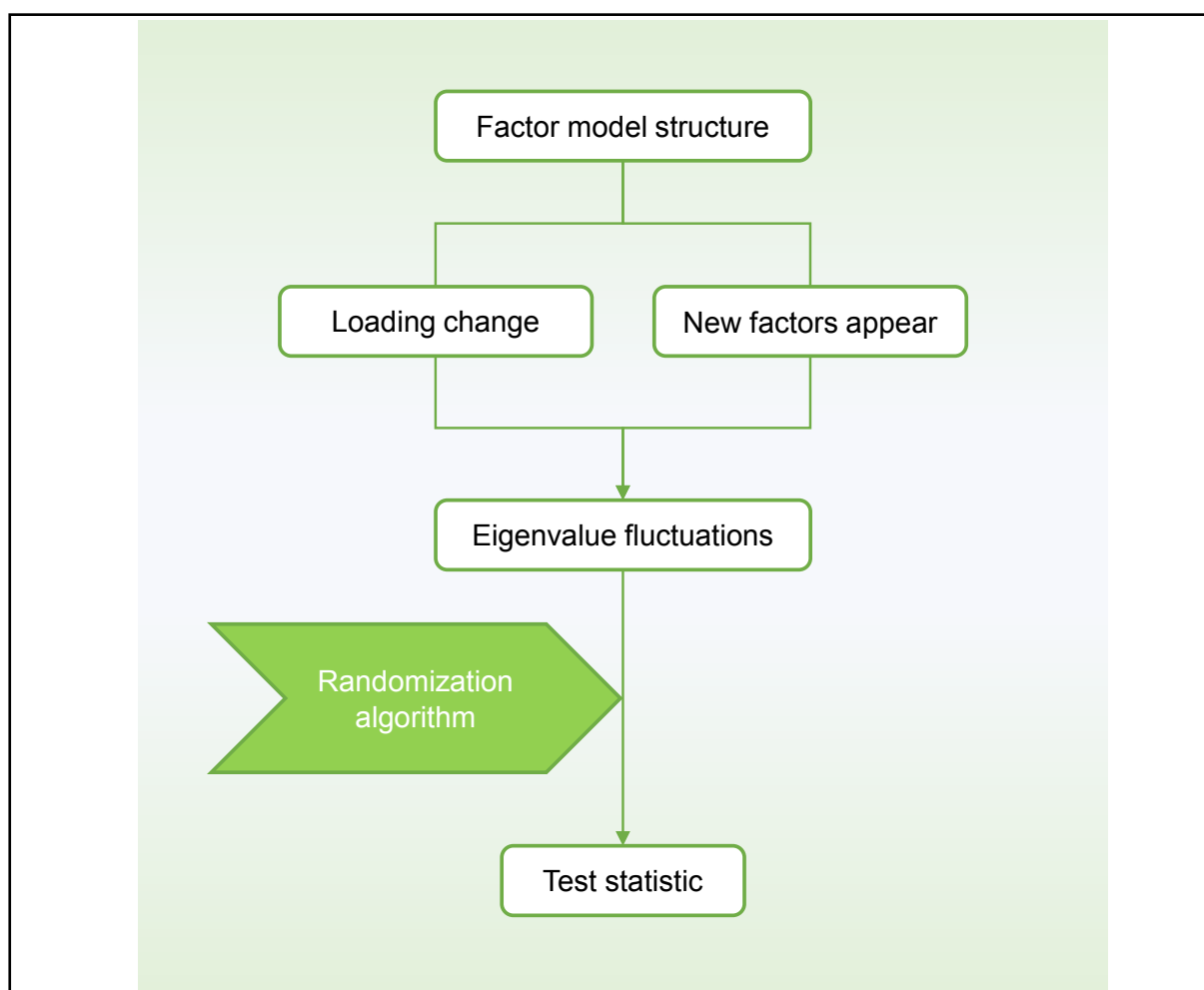
# The fluctuation of eigenvalues in factor models

Fanglin Bao, and Bo Zhang ✉

*Department of Statistics and Finance, International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230026, China*

✉Correspondence: Bo Zhang, E-mail: wbchpmp@ustc.edu.cn

**Abstract:** We consider the fluctuation of eigenvalues in factor models and propose a new method for testing the model. Based on the characteristics of eigenvalues, variables of unknown distribution are transformed into statistics of known distribution through randomization. The test statistic checks for breaks in the structure of factor models, including changes in factor loadings and increases in the number of factors. We give the results of simulation experiments and test the factor structure of the stock return data of China's and U.S. stock markets from January 1, 2017, to December 31, 2019. Our method performs well in both simulations and real data.

**Keywords:** factor models; eigenvalue fluctuation; high-dimensional data; random matrix theory

## 1 Introduction

Currently, people can easily obtain higher-dimensional and larger amounts of data. The factor model, a well-established technique for dimension reduction, is one of the most commonly used methods in high-dimensional research. For the theory related to factor models, see papers such as Refs. [1, 2]. With the wide application of factor models, related theories have also been developed, such as articles related to determining the number of factors in the model[3, 4]. These methods often involve the use of covariance matrices. Let $N$ and $T$ denote the cross-sectional and time series dimensions, respectively. When the dimension $N$ is fixed while $T$ tends to infinity, the eigenvalues of the sample covariance matrix are good estimates of the eigenvalues of the population covariance matrix. However, when both $N$ and $T$ tend to infinity, the estimation at this time is biased. For the study of the eigenvalues of the sample covariance matrix, there is a substantial body of work dealing with the limiting behavior of the eigenvalues. It forms an important subset of what is often called random matrix theory. Related research includes Refs. [5, 6].

In this paper, we consider spiked covariance matrices in the high-dimensional setting, which arise naturally from factor models. A special type of spiked covariance matrix whose nonspiked eigenvalues are equal to 1 is proposed by Ref. [7]. Subsequent studies based on the spiked matrix have been extended in various directions, including the asymptotic distribution of spiked eigenvalues (e.g., Refs. [8–12]). Related studies extend the scope to more generalized matrices, for example, by removing the restriction that all nonspiked eigenvalues equal 1. Based on the eigenvalues of the covariance matrix, Ref. [13] proposes a randomization method and constructs a test statistic to determine the number of factors in a factor model. Applications of this randomization method can be traced back to Ref. [14], and Ref. [15] also applies the method in the financial field. This randomization method is also the main method used in our article.

In a typical factor model, the whole sample is often used to construct a model suitable for the whole period. Time invariance of factor loadings is a standard assumption in the analysis of large factor models. However, such an assumption is often too ideal to suit practice. In reality, the factor structure may constantly change due to the influence of various factors. This study focuses on two types of changes: the instability of factor loadings and the fact that new factors emerge after the change point. Breaks in the factor structure have been explored in many papers. Refs. [16, 17] focus on structural instability in the loadings. Some articles consider the breaks in factor models before and after some significant events. Ref. [18] assesses the evidence for a change in the number of factors during the Great Recession. However, their approach requires prior knowledge of the breaking point and cannot distinguish the two change types we just mentioned. Studies also aim to find the exact time point when a break occurs, which we call a change point, with Ref. [19] being one example.

The problem is that when there is no structural change in the process, some detection methods may give a false result. To avoid such false test results, we need to test whether there are structural changes before conducting in-depth research, and when there are no structural changes, additional studies are unnecessary. In this article, we consider the fluctuation of the eigenvalues in the factor model and use the randomization method to construct a statistic to test whether there is a structural change in the model within the time range. Employing a sliding window, our technique is more sensitive to detecting changes in factor loadings than simply dividing the sample into two. The test statistic performs well in the

absence of change points. The presence of a change point is consistently detected under both types of structural changes (loading instability and new factors).

Our research always assumes that $\sqrt{T}/N = o(1)$. For any matrix $A$, $\|A\|$ is the square root of the largest eigenvalue of $A^\top A$. The rest of the paper is organized as follows. In Section 2, we introduce the model and basic results, and Section 3 presents the theory on which the method is mainly based. Section 4 explains how we construct statistics using the randomization method. Then, in Section 5, simulation experiments are carried out to test the effect of the statistic. In addition, a real data example is reported in Section 6. Finally, we give our conclusions in Section 7. All technical proofs are presented in Appendix.

## 2  The model and basic results

Consider a factor model as follows.

$$x_{it} = \sum_{k=1}^{r} a_{ik} f_{kt} + e_{it}, \ 1 \leqslant i \leqslant N \text{ and } 1 \leqslant t \leqslant T. \tag{1}$$

We can also write the factor model with matrices.

$$X = AF + E, \tag{2}$$

where $X$ and $E$ are two $N \times T$ matrices, $A$ is an $N \times r$ matrix, and $F$ is an $r \times T$ matrix. Define $T^{-1} \sum_{t=1}^{T} f_{kt} = \bar{f}_k$.

**Assumption 1.** Both $N$ and $T$ go to infinity. $A^\top A = I_r$. Moreover, $\lambda_1 \geqslant \cdots \geqslant \lambda_r \asymp N(1 + NT^{-1})$ such that

$$T^{-1} \sum_{t=1}^{T} (f_{kt} - \bar{f}_k)(f_{jt} - \bar{f}_j) = \lambda_j^{1/2} \lambda_k^{1/2} \big(\mathbf{1}\{j = k\} + O_p(T^{-1/2})\big), \tag{3}$$

for any $1 \leqslant j, k \leqslant r$.

Let $M$ be a $T \times T$ matrix that has elements $M_{ij} = \mathbf{1}\{i = j\} - T^{-1}$. Define $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_r)$, then Eq. (3) implies that

$$\left\| \Lambda^{-1/2}(T^{-1} FMF^\top - \Lambda)\Lambda^{-1/2} \right\| = O_p(T^{-1/2}). \tag{4}$$

**Remark 1.** Eq. (3) holds when $f_t = (f_{1t}, \cdots, f_{rt})$ satisfies the following conditions: $f_t$ is strictly stationary and $\psi$-mixing with $\max_{1 \leqslant i \leqslant r} E(\lambda_i^{-2} f_{it}^4) < C < \infty$, $\mathrm{Cov}(f_t) = \Lambda$ and $\sum_{t \geqslant 1} t \psi(t)^{1/2} < \infty$. If the population covariance matrix is not diagonal, it is easy to adjust it by eigendecomposition and a different $A$. However, Eq. (3) may still hold in some other cases even though $f_t$ is a nonrandom time trend. Thus, we impose Assumption 1 on the samples, not the populations, mainly to cover a more extensive range of situations.

**Assumption 2.** $E = \Sigma^{1/2} Y \Omega$. Here, $Y$ has i.i.d. elements $Y_{ij}$ with mean 0, variance 1, and a finite fourth moment. $\Sigma$ is an $N \times N$ nonnegative definition matrix. $\Omega$ is a $\tilde{T} \times T$ matrix with $\tilde{T} \asymp T$. There exists a constant that does not depend on $N$ and $T$ such that $\|\Sigma\| \leqslant C$ and $\|\Omega\| \leqslant C$.

It also implies that $\|T^{-1/2} Y\| = O_p(1 + N^{1/2} T^{-1/2})$ and $\|T^{-1/2} E\| = O_p(1 + N^{1/2} T^{-1/2})$. Let $S = T^{-1} X M X^\top$ be the sample covariance matrix of $X$. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_N \geqslant 0$ be the eigenvalues of $S$.

**Proposition 1.** Under Assumptions 1 and 2,

$$\lambda_{r+1} = O_p(1 + NT^{-1}), \tag{5}$$

and

$$\frac{1 + NT^{-1}}{\lambda_r} = O_p(N^{-1}). \tag{6}$$

Proposition 1 is a simple result, but it provides spectrum separation. Let $U = (A \ B)$ be an $N \times N$ orthogonal matrix. Denote $(F + A^\top E)M$ as $\tilde{F}$ and $B^\top EM$ as $\tilde{E}$.

**Theorem 1.** Under Assumptions 1 and 2,

$$\frac{\lambda_k - \dot{\lambda}_k}{\dot{\lambda}_k} = O_p\big(T^{-1/2} + N^{-1/2}\big), \tag{7}$$

for any $1 \leqslant k \leqslant r$. Moreover, if $\|A^\top E P_F\| = O_p(1 + N^{1/2} T^{-1/2})$, then

$$\frac{\lambda_k - \dot{\lambda}_k}{\dot{\lambda}_k} = O_p\big(T^{-1/2} + N^{-1}\big), \tag{8}$$

for any $1 \leqslant k \leqslant r$. Here, $P_F = MF^\top(FMF^\top)^{-1}FM$.

Theorem 1 shows the fluctuation of $\lambda_k$ from $\dot{\lambda}_k$ for $1 \leqslant k \leqslant r$. Note that when $F$ and $E$ are independent, $A^\top E P_F$ degenerates to a separable model with rank $r$ on both sides. Thus, $\|A^\top E P_F\| = O_p(1 + N^{1/2} T^{-1/2})$ holds. It can also extend to other cases where there is only weak dependence between $F$ and $E$.

**Remark 2.** We note that $\dot{\lambda}_k$ only depends on the population of $F$. Eqs. (7) and (8) imply that the fluctuation of $\lambda_k$ is small even if we replace $E$ by a different population satisfying Assumption 2. Thus, $\lambda_k$ becomes large only if the population of $F$ becomes large.

**Assumption 3.** $N/T \to c < \infty$ as $N$ and $T$ go to infinity, where $c$ is away from 0. Let $\mu_1 \geqslant \cdots \geqslant \mu_{N-r} \geqslant 0$ be the eigenvalues of $T^{-1} B^\top E E^\top B$. There exists a constant $C$ such that

$$\sup_{1 \leqslant k \leqslant K} |\mu_k - C| = O_p(T^{-2/3}), \tag{9}$$

where $K$ is any finite integer.

Eq. (9) is a common result for nonspiked eigenvalues in random matrix theory, and it covers many cases. For example, the asymptotic distribution for the first $K$ eigenvalues of sample covariance matrices with the general population has been stated under some moment constraints in Ref. [20]. Ref. [21] also established the Tracy–Widom asymptotic for the $K$ largest eigenvalues of a general class of random Gram matrices. In fact, $C$ is the right endpoint of the support for LSD of $T^{-1} B^\top E E^\top B$. For more details, one can also refer to Refs. [6, 22, 23].

**Theorem 2.** Under Assumptions 1–3,

$$\sup_{1 \leqslant k \leqslant K} |\lambda_{r+k} - C| = O_p(T^{-2/3}), \tag{10}$$

where $K$ is any finite integer.

Theorem 2 demonstrates that excluding the first $r$ largest eigenvalues, the remaining first $K$ eigenvalues will be located within a small border around the right endpoint $C$. Although Assumption 3 covers the nonspiked cases in random matrix theory, there are other cases for $\tilde{E}$ with so-called

"spiked eigenvalues". Spiked eigenvalues often have a convergence rate of $T^{-1/2}$, which makes them differ from non-spiked eigenvalues. Addressing spiked eigenvalues in random matrix theory is sometimes easier than nonspiked eigenvalues, but the situation is different in our problems. The main reason is that the important terms in our proof of Theorem 2, such as $I - P_{\tilde{F}}$ and $I + T^{-2}\tilde{E}\tilde{F}^\top F_1(T^{-1}\tilde{F}\tilde{F}^\top - \lambda I)^{-1}\tilde{F}\tilde{E}^\top$, may change the spiked structure.

We will consider a special case for $\tilde{E}$ with spiked eigenvalues.

**Assumption 4.** Let $F$ and $E$ be independent. $N \asymp T$. Moreover, $E$ can be rewritten as

$$E = \Sigma^{1/2}Y, \tag{11}$$

where $Y$ has the i.i.d. elements $Y_{ij}$ with mean 0, variance 1, and a finite fourth moment. $\Sigma$ is an $N \times N$ matrix that is not negative definite, with eigenvalues $\infty > \sigma_1 \geqslant \cdots \geqslant \sigma_N \geqslant 0$.

**Theorem 3.** Under Assumptions 1 and 4, the spectrum of $B^\top \Sigma B$ and the distributions of $Y_{ij}$ satisfy the conditions of Theorem 3.6 in Ref. [22]. Then, there exist constants $c_1 \geqslant \cdots \geqslant c_K$ such that

$$\lambda_{r+k} - c_k = O_p(N^{-1}) + O_p(T^{-1/2}), \ 1 \leqslant k \leqslant K. \tag{12}$$

Here, $c_1 \geqslant \cdots \geqslant c_K$ depend only on $\Sigma$.

Assumption 4 gives a situation with spiked eigenvalues in $\tilde{E}$. In this case, (12) establishes a similar result to Theorem 2. The eigenvalues $\lambda_{r+1},\ldots,\lambda_{r+K}$ will still fluctuate in a small range near the limits.

## 3 Fluctuation of eigenvalues

Now, we consider $x_{it}$ with $1 \leqslant t \leqslant T^*$. Define the $N \times T$ matrix $X^{(s)}$ with elements $x_{it}$ $(1 \leqslant i \leqslant N, s \leqslant t \leqslant s + T - 1)$ and $X^{(s)} = A^{(s)}F^{(s)} + E^{(s)}$. For each $s$, we can calculate the $r + 1$ largest eigenvalues of the sample covariance matrix for $X^{(s)}$. We use $\lambda_k^{(s)}$ to represent the $k$th eigenvalue of $X^{(s)}$.

We do not impose any conditions on the correlations between $X^{(s)}$ for varying $s$. In contrast, we want to study whether differences exist between the $X^{(s)}$ structures based on the eigenvalues of their sample covariance matrices. We define $S^{(s)}$ and $\lambda_1^{(s)} \geqslant \cdots \geqslant \lambda_N^{(s)} \geqslant 0$ by replacing $X$ by $X^{(s)}$ in $S$. Similarly, we can define $\Lambda^{(s)}$ by replacing $F$ by $F^{(s)}$ in $S$. Denote that

$$\bar{\lambda}_k = \frac{1}{T^* - T + 1} \sum_{s=1}^{T^*-T+1} \lambda_k^{(s)}. \tag{13}$$

**Theorem 4.** Suppose that Assumptions 1 and 2 hold for all $X^{(s)}$. Then, from Eq. (7), for any $1 \leqslant k \leqslant r$,

$$\frac{\lambda_k^{(s)} - \dot{\lambda}_k^{(s)}}{\dot{\lambda}_k^{(s)}} = O_p\left(T^{-1/2} + N^{-1/2}\right). \tag{14}$$

If $\Lambda^{(s_1)} = \Lambda^{(s_2)}$ for $\forall s_1, s_2$ with $1 \leqslant s_1, s_2 \leqslant T^* - T + 1$, then

$$\frac{|\lambda_k^{(s)} - \bar{\lambda}_k|}{\bar{\lambda}_k} = O_p(T^{-1/2} + N^{-1/2}), \tag{15}$$

for any $1 \leqslant k \leqslant r$. Moreover, if

$$A^{(s)\top}E^{(s)}P_{F^{(s)}}E^{(s)\top}A^{(s)} = O_p(1 + NT^{-1}), \tag{16}$$

also holds,

$$\frac{|\lambda_k^{(s)} - \bar{\lambda}_k|}{\bar{\lambda}_k} = O_p(T^{-1/2} + N^{-1}), \tag{17}$$

for any $1 \leqslant k \leqslant r$.

**Remark 3.** If $N$ is not large enough, we may consider a faster rate of $\lambda_k^{(s)}$. However, $A^{(s)}$ and $E^{(s)}$ impact $\lambda_k^{(s)}$ in this case. Eq. (17) can have a faster rate provided that there are no differences (in distributions) between $(A^{(s)}, E^{(s)})$.

We also consider an alternative hypothesis and result: Assume $s_1$ and $s_2$ exist s.t. $\Lambda^{(s_1)} \neq \Lambda^{(s_2)}$. Then, there may exist $1 \leqslant k \leqslant r$ such that the left part of Eq. (17) has a larger order than $T^{-1/2} + N^{-1/2}$.

Theorem 4 states that for the first $r$ eigenvalues, the left part of Eq. (15) should be small and not fluctuate greatly when the structure has not changed. It is a natural conclusion drawn from Theorem 1.

**Theorem 5.** Suppose that Assumptions 1–3 hold for all $X^{(s)}$. Then, according to Eq. (10),

$$\sup_{1 \leqslant k \leqslant K} |\lambda_{r+k}^{(s)} - C^{(s)}| = O_p(T^{-2/3}), \tag{18}$$

where $K$ is any finite integer. If $C^{(s_1)} = C^{(s_2)}$ for $\forall s_1, s_2$ with $1 \leqslant s_1, s_2 \leqslant T^* - T + 1$, then

$$\frac{|\lambda_{r+k}^{(s)} - \bar{\lambda}_{r+k}|}{\bar{\lambda}_{r+k}} = O_p(T^{-2/3}), \ 1 \leqslant k \leqslant K. \tag{19}$$

**Theorem 6.** Suppose that the conditions of Theorem 3 hold for all $X^{(s)}$. Then,

$$\lambda_{r+k}^{(s)} - c_k^{(s)} = O_p(N^{-1}) + O_p(T^{-1/2}), \ 1 \leqslant k \leqslant K. \tag{20}$$

For $\forall s_1, s_2$ with $1 \leqslant s_1, s_2 \leqslant T^* - T + 1$, if $\Sigma^{(s_1)}$ and $\Sigma^{(s_2)}$ have the same spectrum, then

$$\frac{|\lambda_{r+k}^{(s)} - \bar{\lambda}_{r+k}|}{\bar{\lambda}_{r+k}} = O_p(N^{-1} + T^{-1/2}), \ 1 \leqslant k \leqslant K. \tag{21}$$

**Remark 4.** With an alternative hypothesis, we have the following: If $C^{(s_1)} \neq C^{(s_2)}$ or $\Sigma^{(s_1)}$ and $\Sigma^{(s_2)}$ have different spectra, then there may exist $k$ such that the left part of (21) has a larger order than $T^{-1/2} + N^{-1}$.

Following Theorems 2 and 3, it is easy to obtain Theorems 5 and 6. We know that the left part of (21) should not fluctuate greatly for both the nonspiked and spiked cases when there is no break in the factor structure.

## 4 Randomization algorithm

In Section 3, we consider only the results of orders. We can use the randomization algorithm in Ref. [13] to construct a new test statistic.

Based on Theorems 1–6, we construct the new statistic using

$$\varphi(s,k) = \frac{|\lambda_k^{(s)} - \bar{\lambda}_k|}{\bar{\lambda}_k}, \ \phi = \max_{1 \leqslant s \leqslant T^*-T+1, 1 \leqslant k \leqslant r_{\max}} \varphi(s,k). \tag{22}$$

Here, $r_{\max}$ is a value given in advance. It controls the number of possible factors in all sliding windows and is predetermined in the algorithm by a simple initial estimate: For $X^{(s)}$ in each sliding window, compute $r_s = \arg\max_{1 \leqslant k \leqslant 19} \lambda_k^{(s)}/\lambda_{k+1}^{(s)}$ to

estimate the number of factors for that window. When corresponding estimates are made for all windows, $r_{\max}$ is chosen to be slightly larger than the maximum of $\{r_s\}_{s=1}^{T^*-T+1}$ so that it can always include the $r+1$ largest eigenvalues:

$$r_{\max} = \max_{1 \leqslant s \leqslant T^*-T+1} r_s + 1. \tag{23}$$

To detect whether there is a change point in $1 \leqslant t \leqslant T^*$, the null hypothesis and the alternative hypothesis can be set as follows:

$$\begin{cases} H_0 : \phi = O_p(T^{-1/2}); \\ H_1 : \phi^{-1} = O_p(1). \end{cases} \tag{24}$$

Here, $\phi$ is defined as Eq. (22). Then, the new statistic is constructed by

$$\Phi = \exp\{1/\phi\}. \tag{25}$$

Since $\phi$ has an order of $T^{-1/2}$ under the null hypothesis, then $\Phi$ should diverge to infinity at a rate of $\exp\{T^{1/2}\}$. The randomization algorithm follows as in Ref. [24]:

**Step 1.** For $\Phi$, generate an i.i.d. sample $\{\xi_j\}_{j=1}^R$ with common distribution $G$ (the standard normal distribution).

**Step 2.** For any $u$ drawn from a distribution $F(u)$ with support $U \subset R/\{0\}$, define

$$\zeta_j(u) = I\{\xi_j \times \Phi \leqslant u\}. \tag{26}$$

**Step 3.** Compute

$$\vartheta(u) = \frac{1}{\sqrt{R}} \sum_{j=1}^R \frac{\zeta_j(u) - G(0)}{\sqrt{G(0)[1-G(0)]}}. \tag{27}$$

**Step 4.** Compute

$$\Theta = \int_{-\infty}^{+\infty} |\vartheta(u)|^2 \mathrm{d}F(u). \tag{28}$$

The considerations in Ref. [24] point toward choosing $u = \pm\sqrt{2}$. A relevant explanation of how the test statistics work under the null and alternative hypotheses is as follows. Under the null hypothesis, $\Phi$ will approach infinity as mentioned before; therefore, the i.i.d. sequence $\{\zeta_j(u)\}_{j=1}^R$ follows a Bernoulli distribution with $E[\zeta_j(u)] = G(0)$. Then in Step 3, $\vartheta(u)$ should be $N(0,1)$ because the central limit theorem should hold when $R$ is large enough. Consequently, $\Theta$ in Step 4 should follow a chi-square distribution with one degree of freedom. Under the alternative hypothesis, $\Phi$ remains finite, so for any $u \neq 0$, $E[\zeta_j(u)] \neq G(0)$. Naturally, a sum of i.i.d. random variables with nonzero means is obtained in Step 3. Hence, $\Theta$ also diverges to infinity.

## 5 Simulation

To verify the effect of the constructed statistic, repeated experiments were carried out under each group of conditions, where $r$ represents the number of factors under the null hypothesis. Under $H_0$, we simulate data $x_{it}$ with $1 \leqslant i \leqslant N, 1 \leqslant t \leqslant T^*$ according to the model:

$$x_{it} = a_i^\top f_t + u_{it}, \quad 1 \leqslant i \leqslant N, \ 1 \leqslant t \leqslant T^*, \tag{29}$$

where $T^*$ means the whole process of data monitoring, and $T$ is the chosen window width. We fix $N = 100$ and $T^* = 800$ for the whole process and set the change point $\tau = 500$. Additionally, we consider $r \in \{1,2,3,4\}$ factors and sliding window width $T \in \{150, 200, 250, 300\}$.

Common factors are allowed to have some time dependence:

$$f_t = 0.7 I_r f_{t-1} + e_t, \quad 1 \leqslant t \leqslant T^*, \tag{30}$$

with $e_t \overset{\text{i.i.d.}}{\sim} N(0, I_r)$. We simulate each element of $a_i$ as i.i.d. $N(0,1)$. The idiosyncratic components $u = D\varepsilon G$ with each element of $\varepsilon$ i.i.d. $N(0,1)$. $D$ and $G$ are two Toeplitz matrices of size $N \times N$ and $T^* \times T^*$, respectively. The entries in the $k$th diagonal place are given by $0.3^{k-1}$ for $D$ and $0.5^{k-1}$ for $G$. $R = N$ is always chosen in the randomization step because the construction of the final statistic is based on the central limit theorem. The distribution $G$ in Step 1 is set to be $N(0,1)$, while $F$ in Step 2 is chosen to have nonzero and equal mass at $\pm\sqrt{2}$.

When the factor loadings have changed after the change point, the factor model can be expressed in the following form:

$$x_{it} = a_i^\top f_t \, I[t < \tau] + \widetilde{a}_i^\top f_t \, I[t \geqslant \tau] + u_{it}, \quad 1 \leqslant i \leqslant N, \ 1 \leqslant t \leqslant T^*, \tag{31}$$

where each element of $a$ and $\widetilde{a} \overset{\text{i.i.d.}}{\sim} N(0,1)$.

When using the traditional way to look at two samples separately, the eigenvalues reflect only the information of each one; they cannot reflect the change in loading. However, putting two samples together and calculating through the sliding window method can ensure the eigenvalues reflect that change. One of the examples is that when the common factors stay constant in (29), but there is such a relationship between the factor loadings $A_1 = PA_2$, $P$ is an orthogonal matrix. If we simply study the eigenvalues of the two covariance matrices, we find that they have almost the same eigenvalues, but at this time, the two loading matrices are obviously not equal. At this point, our sliding window method can identify such changes very well.

Consider the case where $r_1$ new factors appear after the change point, and the new factors satisfy: $g_t = 0.7 I_{r_1} g_{t-1} + v_t$, $1 \leqslant t \leqslant T^*$, with $v_t \overset{\text{i.i.d.}}{\sim} N(0, I_{r_1})$. The factor model at this time can be written as follows:

$$x_{it} = a_i^\top f_t + b_i^\top g_t \, I[t \geqslant \tau] + u_{it}, \quad 1 \leqslant i \leqslant N, \ 1 \leqslant t \leqslant T^*. \tag{32}$$

Elements of $b_i$ are generated as i.i.d. $N(0,1)$.

To evaluate the performance of our statistic, we repeat the simulation 300 times under varying conditions and present the outcomes in Tables 1–4. Table 1 reports the fraction of false rejections under $H_0$ (i.e., when there is no break in the factor model) at the 1% and 5% significance level. In simulations, the sample size $T^* = 800$, and the window width $T \in \{150, 200, 250, 300\}$.

Under the null hypothesis, it can be seen from the data in Table 1 that the constructed test statistic can reach a good confidence level when a large window width $T$ is chosen. Moreover, the performance of the test statistic is not affected

**Table 1.** False rejection at 1% and 5% significance ($T^* = 800$).

| | Fraction of false rejections | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | $T = 150$ | | $T = 200$ | | $T = 250$ | | $T = 300$ | |
| $r$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| 1 | 0.049 | 0.158 | 0.022 | 0.078 | 0.016 | 0.062 | 0.011 | 0.047 |
| 2 | 0.042 | 0.204 | 0.004 | 0.058 | 0.009 | 0.053 | 0.007 | 0.036 |
| 3 | 0.029 | 0.138 | 0.007 | 0.069 | 0.004 | 0.056 | 0.009 | 0.040 |
| 4 | 0.027 | 0.129 | 0.011 | 0.069 | 0.007 | 0.027 | 0.013 | 0.038 |

**Table 2.** Power of the statistic at 1% and 5% significance ($T^* = 800$).

| | | Fraction of true rejections | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | $T = 150$ | | $T = 200$ | | $T = 250$ | | $T = 300$ | |
| | $r$ | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| Loadings change | 1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| A new factor appears | 1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.993 | 1.000 |
| Two new factors appear | 1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Table 3.** False rejection at 1% and 5% significance ($T^* = 200$).

| | Fraction of false rejections | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | $T = 100$ | | $T = 125$ | | $T = 150$ | |
| $r$ | 1% | 5% | 1% | 5% | 1% | 5% |
| 1 | 0.010 | 0.087 | 0.017 | 0.050 | 0.013 | 0.063 |
| 2 | 0.020 | 0.097 | 0.020 | 0.057 | 0.007 | 0.043 |
| 3 | 0.013 | 0.070 | 0.023 | 0.057 | 0.023 | 0.070 |
| 4 | 0.013 | 0.073 | 0.003 | 0.033 | 0.007 | 0.057 |

by increases in the number of factors in the original model.

In Table 2, we show the fraction of detections for three types of structural changes when a break takes place and when testing at 1% and 5% significance levels. Under the alternative hypothesis, considering the changes in the factor loadings and the number of factors, the statistic can reject the null hypothesis with a probability of 1, which indicates that it can test the changes in the structure very well. At the same time, the results in Table 2 show that it can have strong power under the alternative assumptions of different factor numbers in the original model and newly added factor numbers.

Considering a smaller sample size $T^* = 200$ with change point $\tau = 100$, Table 3 shows the false rejection rate under $H_0$ of selecting the window width $T \in \{100, 125, 150\}$. According to the results in the table, even if $T^*$ is reduced to 200, our

method can still be used when window width $T = 125$. This suggests that a larger window width can bring a better size of the test. However, it is not appropriate to select a window width $T$ that is too large for the whole sample size $T^*$. Table 4 shows the fraction of detections under the alternative hypothesis when testing at 1% and 5% significance levels. Alghouth Table 3 shows that it has a better size with window width $T = 150$, the power in Table 4 is too low for the test.

## 6 Real data

Taking the daily returns of the S&P 500 and CSI 300 constituent stocks in China's and U.S. stock markets, respectively, from January 1, 2017, to December 31, 2019, as the research object, we examine the factor structure of the stock

**Table 4.** Power of the statistic at 1% and 5% significance ($T^* = 200$).

| | | Fraction of true rejections | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | $T = 100$ | | $T = 125$ | | $T = 150$ | |
| | $r$ | 1% | 5% | 1% | 5% | 1% | 5% |
| Loadings change | 1 | 0.800 | 0.957 | 0.153 | 0.430 | 0.013 | 0.063 |
| | 2 | 0.860 | 0.977 | 0.173 | 0.483 | 0.013 | 0.073 |
| | 3 | 0.887 | 0.980 | 0.153 | 0.510 | 0.007 | 0.050 |
| | 4 | 0.897 | 0.990 | 0.133 | 0.450 | 0.010 | 0.057 |
| A new factor appears | 1 | 0.943 | 1.000 | 0.347 | 0.597 | 0.037 | 0.123 |
| | 2 | 0.857 | 0.983 | 0.240 | 0.543 | 0.033 | 0.123 |
| | 3 | 0.790 | 0.977 | 0.277 | 0.530 | 0.027 | 0.107 |
| | 4 | 0.823 | 0.990 | 0.193 | 0.473 | 0.013 | 0.060 |
| Two new factors appear | 1 | 0.990 | 1.000 | 0.520 | 0.840 | 0.033 | 0.147 |
| | 2 | 0.957 | 0.997 | 0.413 | 0.777 | 0.053 | 0.180 |
| | 3 | 0.953 | 0.997 | 0.313 | 0.683 | 0.040 | 0.157 |
| | 4 | 0.903 | 0.993 | 0.337 | 0.640 | 0.027 | 0.123 |

markets of the two countries. Stocks that have been suspended for more than 60 consecutive days are eliminated, and the remaining missing stock return data are filled with 0. Aggregating the data according to the same opening date for both markets, we obtain the daily returns of 749 stocks for 708 trading days.

We use the randomization algorithm in Section 4 to test the factor structure of the above data. The window width is selected as $T = 200$, and the final test statistic value is $\Theta = 58.94$, which exceeds the critical value $c_\alpha = 6.635$ when the significance level is $\alpha = 0.01$. This suggests that there are changes in the factor structure during this period. To show the result more clearly, Fig. 1 presents the value of $\lambda_1^{(s)}$ in each sliding window $\boldsymbol{X}^{(s)}$. In the whole process, $\lambda_1^{(s)}$ changes greatly with the movement of the window, which also shows that the factor structure has changed.

The selected data include the period of the U.S.-China trade war, during which many important trade war events occurred, including multiple mutual taxation and meeting negotiations between China and the United States. The trade relationship between these two countries has experienced huge volatility, and stock markets have also fluctuated accordingly. Hence, the test result is consistent with the actual situation.

# 7 Conclusions

We study the fluctuation of eigenvalues in factor models and construct a test statistic of the model structure. One only need to detect the change points after the test shows the existence of change points. It reduces unnecessary work. We also apply our method to simulations and the real data from China's and U.S. stock markets. Our method performs well in both simulations and real data.

At the same time, the sliding window method also avoids the insensitivity of the method of simply comparing two samples for eigenvalue research to determine the change in factor loadings. Our proposed test statistic performs well under both types of structural changes (the instability of
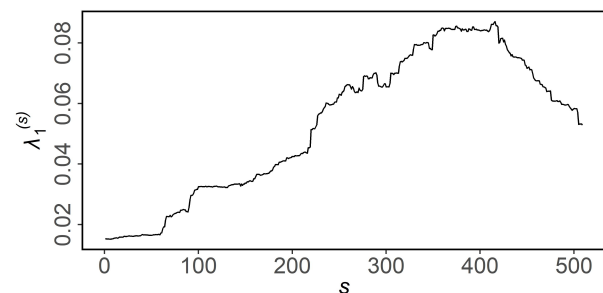


**Fig. 1.** The largest eigenvalue $\lambda_1^{(s)}$ in each sliding window (for $1 \leqslant s \leqslant 509$).

factor loadings and the fact that new factors emerge after the change point) and is not affected by the number of factors in the original model. Accordingly, this novel statistic has good universality for application in future studies.

# Conflict of interest

The authors declare that they have no conflict of interest.

# Biographies

**Fanglin Bao** is a graduate student under the tutelage of Prof. Bo Zhang at the University of Science and Technology of China. Her research mainly focuses on high-dimensional statistics.

**Bo Zhang**  is an Associate Professor at the University of Science and Technology of China (USTC). He obtained a Bachelor's degree in Statistics from USTC in 2011, a Master's degree in Probability Theory and Mathematical Statistics from USTC in 2013, and a Ph.D. degree in Mathematics from Nanyang Technological University in Singapore in 2017. His research mainly focuses on high-dimensional random matrices, high-dimensional time series, and complex networks.

## References

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

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

[3] Bai J, Ng S. Determining the number of factors in approximate factor models. *Econometrica,* **2002**, *70* (1): 191–221.

[4] Onatski A. Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics,* **2010**, *92* (4): 1004–1016.

[5] El Karoui N. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics,* **2008**, *36* (6): 2757–2790.

[6] Bao Z, Pan G, Zhou W. Universality for the largest eigenvalue of sample covariance matrices with general population. *The Annals of Statistics,* **2015**, *43* (1): 382–421.

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

[8] Baik J, Silverstein J W. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis,* **2006**, *97* (6): 1382–1408.

[9] Paul D. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica,* **2007**, *17* (4): 1617–1642.

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

[11] Wang W, Fan J. Asymptotics of empirical eigenstructure for high

[12] dimensional spiked covariance. *The Annals of Statistics,* **2017**, *45* (3): 1342–1374.

[12] Cai T T, Han X, Pan G. Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *The Annals of Statistics,* **2020**, *48* (3): 1255–1280.

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

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

[15] Corradi V, Swanson N R. The effect of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo results and a simple test. *Journal of Econometrics,* **2006**, *132* (1): 195–229.

[16] Chen L, Dolado J J, Gonzalo J. Detecting big structural breaks in large factor models. *Journal of Econometrics,* **2014**, *180* (1): 30–48.

[17] Cheng X, Liao Z, Schorfheide F. Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies,* **2016**, *83* (4): 1511–1543.

[18] Stock J H, Watson M W. Disentangling the channels of the 2007–2009 recession. Cambridge, MA: National Bureau of Economic Research, **2012**: 18094.

[19] Breitung J, Eickmeier S. Testing for structural breaks in dynamic factor models. *Journal of Econometrics,* **2011**, *163* (1): 71–84.

[20] Ding X, Yang F. A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *The Annals of Applied Probability,* **2018**, *28* (3): 1679–1738.

[21] Ding X, Yang F. Tracy–Widom distribution for heterogeneous Gram matrices with applications in signal detection. *IEEE Transactions on Information Theory,* **2022**, *68* (10): 6682–6715.

[22] Ding X, Yang F. Spiked separable covariance matrices and principal components. *The Annals of Statistics,* **2021**, *49* (2): 1113–1138.

[23] Knowles A, Yin J. Anisotropic local laws for random matrices. *Probability Theory and Related Fields,* **2017**, *169* (1): 257–352.

[24] Barigozzi M, Trapani L. Sequential testing for structural stability in approximate factor models. *Stochastic Processes and Their Applications,* **2020**, *130* (8): 5149–5187.

# Appendix

## A.1   Proof of Theorem 1

Let the eigenvalues of $U^\top S U$ be $\lambda_1 \geqslant \cdots \geqslant \lambda_N \geqslant 0$. Note that $M^2 = M$.

$$U^\top S U = T^{-1} \begin{pmatrix} A^\top S A & A^\top S B \\ B^\top S A & B^\top S B \end{pmatrix} = T^{-1} \begin{pmatrix} \tilde{F}\tilde{F}^\top & \tilde{F}\tilde{E}^\top \\ \tilde{E}\tilde{F}^\top & \tilde{E}\tilde{E}^\top \end{pmatrix}. \tag{A1}$$

Recalling that $\|T^{-1/2}E\| = O_p(1 + N^{1/2}T^{-1/2})$, $\|\tilde{E}\tilde{E}^\top\| \leqslant \|E\|^2$, and Eq. (6), we can find that $\lambda_i I - T^{-1}\tilde{E}\tilde{E}^\top$ is positive-definite with probability 1 (as $N$, and $T$ tends to infinity) for any $i \leqslant r$. Then, for any $i \leqslant r$, $\lambda_i$ is the solution of

$$\det\left(\lambda \Lambda^{-1} - T^{-1}\Lambda^{-1/2}\tilde{F}\tilde{F}^\top\Lambda^{-1/2} - T^{-2}\lambda^{-1}\Lambda^{-1/2}\tilde{F}\tilde{E}^\top(I - \lambda^{-1}T^{-1}\tilde{E}\tilde{E}^\top)^{-1}\tilde{E}\tilde{F}^\top\Lambda^{-1/2}\right) = 0, \tag{A2}$$

with probability 1 (as $N$ and $T$ tend to infinity). Similarly, we can find that $\|\lambda^{-1}T^{-1}\tilde{E}\tilde{E}^\top\| = O_p(N^{-1})$ when $\lambda$ is a solution of Eq. (A2). Then, we can rewrite the last term of the left part of Eq. (A2) as

$$\sum_{s=0}^{s_1} T^{-2-s}\lambda^{-1-s}\Lambda^{-1/2}\tilde{F}\tilde{E}^\top(\tilde{E}\tilde{E}^\top)^s\tilde{E}\tilde{F}^\top\Lambda^{-1/2} + O_p(N^{-(2+s_1)}) = O_p(N^{-1}). \tag{A3}$$

Now, we consider the second term of the left part in Eq. (A2). We rewrite $\tilde{F}\tilde{F}^\top$ as follows:

$$\tilde{F}\tilde{F}^{\top} = FMF^{\top} + FME^{\top}A + A^{\top}EMF^{\top} + A^{\top}EME^{\top}A. \tag{A4}$$

Recalling Eqs. (3) and (4), $\|FMF^{\top}\|_{\min}$ has order $TN(1 + NT^{-1})$. This, together with the upper bound of $E$, implies that,

$$\left\|(FMF^{\top})^{-1/2}A^{\top}EMF^{\top}(FMF^{\top})^{-1/2}\right\| \leqslant \|FMF^{\top}\|_{\min}^{-1/2}\|E\|\left\|MF^{\top}(FMF^{\top})^{-1/2}\right\| = O_p(N^{-1/2}),$$

and

$$\left\|(FMF^{\top})^{-1/2}A^{\top}EME^{\top}A(FMF^{\top})^{-1/2}\right\| = O_p(N^{-1}).$$

Then

$$\tilde{F}\tilde{F}^{\top} = FMF^{\top}\left(1 + O_p(N^{-1/2}) + O_p(N^{-1})\right).$$

This, together with Eqs. (4), (A2), and (A3), concludes the proof of Eq. (7).

Moreover, if $\|A^{\top}EP_F\| = O_p(1 + N^{1/2}T^{-1/2})$, then

$$\left\|A^{\top}EMF^{\top}(FMF^{\top})^{-1/2}\right\| = \|A^{\top}EP_FE^{\top}A\|^{1/2} = O_p(1 + N^{1/2}T^{-1/2}).$$

Thus,

$$\left\|(FMF^{\top})^{-1/2}A^{\top}EMF^{\top}(FMF^{\top})^{-1/2}\right\| \leqslant \|A^{\top}EP_F\|^{1/2}\left\|(FMF^{\top})^{-1/2}\right\| = O_p(T^{-1/2}N^{-1/2}),$$

and

$$\left\|(FMF^{\top})^{-1/2}FME^{\top}A(FMF^{\top})^{-1/2}\right\| \leqslant \|A^{\top}EP_F\|^{1/2}\left\|(FMF^{\top})^{-1/2}\right\| = O_p(T^{-1/2}N^{-1/2}).$$

Then, $\tilde{F}\tilde{F}^{\top} = FMF^{\top}\left(1 + O_p(T^{-1/2}N^{-1/2}) + O_p(N^{-1})\right)$.

This, together with Eqs. (4), (A2), and (A3), concludes the proof of Eq. (8).

## A.2 Proof of Theorem 2

For any $i > r$, $T^{-1}\tilde{F}\tilde{F}^{\top} - \lambda_i I$ is positive-definite with probability 1 (as $N$ and $T$ tend to infinity). Then, for any $i > r$, $\lambda_i$ is the solution of

$$\det\left(\lambda I - T^{-1}\tilde{E}\tilde{E}^{\top} - T^{-2}\tilde{E}\tilde{F}^{\top}(\lambda I - T^{-1}\tilde{F}\tilde{F}^{\top})^{-1}\tilde{F}\tilde{E}^{\top}\right) = 0. \tag{A5}$$

Define $P_{\tilde{F}} = \tilde{F}^{\top}(\tilde{F}\tilde{F}^{\top})^{-1}\tilde{F}$ and $F_1 = (T^{-1}\tilde{F}\tilde{F}^{\top})^{-1}$. It is easy to see that $\|F_1\| = O_p(N^{-1}\lambda^{-1})$ when $\lambda$ is the solution of Eq. (A5).

$$(\lambda I - T^{-1}\tilde{F}\tilde{F}^{\top})^{-1} = -F_1 + \lambda F_1(\lambda I - T^{-1}\tilde{F}\tilde{F}^{\top})^{-1}. \tag{A6}$$

Then, Eq. (A5) can be rewritten as

$$\det\left(\lambda(I + T^{-2}\tilde{E}\tilde{F}^{\top}F_1(T^{-1}\tilde{F}\tilde{F}^{\top} - \lambda I)^{-1}\tilde{F}\tilde{E}^{\top}) - T^{-1}\tilde{E}(I - P_{\tilde{F}})\tilde{E}^{\top}\right) = 0. \tag{A7}$$

First, let us consider the $k$ largest eigenvalues of $T^{-1}\tilde{E}(I - P_{\tilde{F}})\tilde{E}^{\top}$. Since $\mu_1 \geqslant \cdots \geqslant \mu_{N-r} \geqslant 0$ are the eigenvalues of $T^{-1}B^{\top}EE^{\top}B$, we can run SVD on $T^{-1/2}B^{\top}E$. Then we can rewrite $T^{-1/2}B^{\top}E$ as

$$T^{-1/2}B^{\top}E = \sum_{i=1}^{N-r}\mu_i^{1/2}u_i v_i^{\top}.$$

Note that

$$T^{-1}\tilde{E}(I - P_{\tilde{F}})\tilde{E}^{\top} = T^{-1}B^{\top}EM(I - P_{\tilde{F}})ME^{\top}B,$$

and $M(I - P_{\tilde{F}})M$ is not a negative-definite matrix, and the largest eigenvalue of $I - P_{\tilde{F}}$ is 1. Then, the $k$ largest eigenvalues of $T^{-1}\tilde{E}(I - P_{\tilde{F}})\tilde{E}^{\top}$ are not larger than $\mu_1, \cdots, \mu_k$ but also not smaller than the first $k$ largest eigenvalues of

$$M(I - P_{\tilde{F}})M\sum_{i=1}^{K}\mu_i v_i v_i^{\top} = M(I - P_{\tilde{F}})MV_K\Lambda_K V_K^{\top}, \tag{A8}$$

where $V_K = (v_1, \cdots, v_K)$, $\Lambda_K = \text{diag}\{\mu_1, \cdots, \mu_K\}$, and $K > k + r$.

Note that $P_{\tilde{F}}$ is a projection matrix with rank $r$. Therefore, $V_K^{\top}M(I - P_{\tilde{F}})MV_K$ has at least $K - r - 1$ eigenvalues 1. This, together with Eqs. (9) and (A8), implies that the $k$ largest eigenvalues of $T^{-1}\tilde{E}(I - P_{\tilde{F}})\tilde{E}^{\top}$ are $C + O_p(T^{-2/3})$.

Similarly, we can repeat the above idea on

$$(I + T^{-2}\tilde{E}\tilde{F}^{\top}F_1(T^{-1}\tilde{F}\tilde{F}^{\top} - \lambda I)^{-1}\tilde{F}\tilde{E}^{\top})^{-1}T^{-1}\tilde{E}(I - P_{\tilde{F}})\tilde{E}^{\top}.$$

We find that $I + T^{-2}\tilde{E}\tilde{F}^{\top}F_1(T^{-1}\tilde{F}\tilde{F}^{\top} - \lambda I)^{-1}\tilde{F}\tilde{E}^{\top}$ is positive-definite with probability 1, and the rank of $T^{-2}\tilde{E}\tilde{F}^{\top}F_1(T^{-1}\tilde{F}\tilde{F}^{\top} - \lambda I)^{-1}\tilde{F}\tilde{E}^{\top}$ is not larger than $r$. Moreover,

$$\left\| I + T^{-2}\tilde{E}\tilde{F}^{\top}F_1(T^{-1}\tilde{F}\tilde{F}^{\top} - \lambda I)^{-1}\tilde{F}\tilde{E}^{\top} \right\|_{\min} \geqslant 1.$$

Thus,

$$\left\| (I + T^{-2}\tilde{E}\tilde{F}^{\top}F_1(T^{-1}\tilde{F}\tilde{F}^{\top} - \lambda I)^{-1}\tilde{F}\tilde{E}^{\top})^{-1} \right\| = 1,$$

and it has at least $N - 2r$ eigenvalues of 1. This lets us conclude Eq. (10).

## A.3    Proof of Theorem 3

Let $\tilde{\lambda}_{0,r+1} \geqslant \cdots \geqslant \tilde{\lambda}_{0,N} \geqslant 0$ be the eigenvalues of

$$T^{-1}\tilde{E}(I - P_F)\tilde{E}^{\top} = T^{-1}B^{\top}\Sigma^{1/2}YM(I - P_F)MY^{\top}\Sigma^{1/2}B. \tag{A9}$$

We know that $M(I - P_F)M$ is independent of $E$ since $F$ and $E$ are independent. Eq. (A.9) can be considered a separable sample covariance matrix when we freeze $F$. $M(I - P_F)M$ has at least $T - r - 1$ eigenvalues 1 and $r$ eigenvalues 0. This, together with Assumption 4, satisfies the conditions of Theorem 3.6 in Ref. [22]. Recalling Theorem 3.6 in Ref. [22], the spiked eigenvalues have a rate of $T^{-1/2}$ and the nonspiked eigenvalues have a rate of $T^{-2/3}$. Since $T^{-2/3} = o(T^{-1/2})$, we find that the fluctuations of the $K$ largest eigenvalues are $O_p(T^{-1/2})$. Thus, we can prove that there exist constants $c_1 \geqslant \cdots \geqslant c_K$ such that

$$\tilde{\lambda}_{0,r+k} - c_k = O_p(T^{-1/2}), \ 1 \leqslant k \leqslant K.$$

Here, $c_1 \geqslant \cdots \geqslant c_K$ only depend on $\Sigma$.

Now, we only need to prove $\lambda_{r+k} - \tilde{\lambda}_{0,r+k} = O_p(N^{-1} + T^{-1/2})$ for $1 \leqslant k \leqslant K$. Recalling Eq. (A7), we need to prove

$$\left\| T^{-2}\tilde{E}\tilde{F}^{\top}F_1(T^{-1}\tilde{F}\tilde{F}^{\top} - \lambda I)^{-1}\tilde{F}\tilde{E}^{\top} \right\| = O_p(N^{-1} + T^{-1/2}), \tag{A10}$$

and

$$P_{\tilde{F}} = P_F(1 + O_p(N^{-1}) + O_p(T^{-1/2})). \tag{A11}$$

Eq. (A10) can be proved by combining the orders of $(T^{-1}\tilde{F}\tilde{F}^{\top} - \lambda I)^{-1}$ and $\tilde{E}\tilde{E}^{\top}$. To prove Eq. (A11), we recall Eq. (A4),

$$\tilde{F}\tilde{F}^{\top} = FMF^{\top} + FME^{\top}A + A^{\top}EMF^{\top} + A^{\top}EME^{\top}A.$$

Recalling Eqs. (3) and (4), $\|FMF^{\top}\|$ has the order of $TN(1 + NT^{-1})$. Note that $F$ and $E$ are independent, so

$$A^{\top}EMF^{\top}(FMF^{\top})^{-1}FME^{\top}A = O_p(1),$$

and

$$A^{\top}EME^{\top}A = O_p(T).$$

Then,

$$\tilde{F}\tilde{F}^{\top} = FMF^{\top}\left(1 + o_p(T^{-1/2}) + O_p(N^{-1})\right),$$

and

$$\tilde{F}^{\top}(\tilde{F}\tilde{F}^{\top})^{-1}\tilde{F} = \tilde{F}^{\top}(FMF^{\top})^{-1}\tilde{F}\left(1 + o_p(T^{-1/2}) + O_p(N^{-1})\right).$$

Note that

$$\tilde{F}^{\top}(FMF^{\top})^{-1}\tilde{F} = MF^{\top}(FMF^{\top})^{-1}FM + ME^{\top}A(FMF^{\top})^{-1}FM + MF^{\top}(FMF^{\top})^{-1}A^{\top}EM + ME^{\top}A(FMF^{\top})^{-1}A^{\top}EM = $$
$$P_F(1 + O_p(N^{-1}) + o_p(T^{-1/2})).$$

Then, $\tilde{F}^{\top}(\tilde{F}\tilde{F}^{\top})^{-1}\tilde{F} = P_F(1 + O_p(N^{-1}) + o_p(T^{-1/2}))$.

This completes the proof.