Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Tests for Principal Eigenvalues and Eigenvectors

Jianqing Fan, Yingying Li, Ningning Xia & Xinghua Zheng

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2026, VOL. 00, NO. 0, 1–11: Theory and Methods
https://doi.org/10.1080/01621459.2026.2635769

Tests for Principal Eigenvalues and Eigenvectors

Jianqing Fan[a#], Yingying Li[b], Ningning Xia[c], and Xinghua Zheng[b]

[a]School of Data Science, Fudan University, Shanghai, China; [b]Department of Information System, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Kowloon, Hong Kong; [c]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China

ABSTRACT

We establish central limit theorems for principal eigenvalues and eigenvectors under divergent spiked covariance models, and develop three two-sample tests for testing (a) the equality of principal eigenvalues, (b) the stability of eigenvalue proportions, and (c) the invariance of principal eigenvectors. One important application of our results is to understand structural breaks in large factor models. That is, when a structural break is suspected or detected, these tests provide unique insights into the source, determining whether the structural break stems from altered factor strength (eigenvalues), a shift in the factors' relative importance (eigenvalue proportions), or a fundamental factor reorientation (eigenvectors). We demonstrate the application by analyzing daily returns of S&P 500 stocks, revealing that major historical structural breaks, such as the 2008 financial crisis and the 2020 pandemic, were complex events characterized by simultaneous shifts in factor strength, relative importance, and orientation. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received May 2024
Accepted February 2026

KEYWORDS
Divergent spiked covariance model; Principal eigenvalue; Principal eigenvector; Structural break; Two-sample test

1. Introduction

The spiked covariance model, introduced in Johnstone (2001), provides a pivotal framework for analyzing high-dimensional data. Its core principle is to separate the total variance into two distinct components: a few "spikes" and a "bulk" of noise. Each spike, along with its corresponding eigenvector, represents a significant source of systematic variation or a strong underlying factor. For instance, in finance, these spikes could correspond to market-wide movements, while in genetics, they might represent the influence of specific ancestry groups. Recent research works on the spiked model include Wang and Fan (2017), Cai, Han, and Pan (2020), Fan, Wang, and Zhu (2021), Bao et al. (2022), Zhang, Cai, and Wu (2022), Yan et al. (2024), Li et al. (2024), among others. In particular, Zhang, Cai, and Wu (2022) and Yan et al. (2024) investigate respectively the estimation and inference of heteroscedastic PCA under the nearly minimum signal-to-noise ratios, but it is studied under the cross-sectionally independent idiosyncratic noise assumption. Recently, Fan, Yan, and Zheng (2024) develop new technical tools to remove the independent idiosyncratic noise, but requires sub-Gaussian assumption. Robustification of PCA has been studied by Fan, Wang, and Zhu (2021) and Wang and Fan (2025) under a more general setting such as the matrix completion, but they also assume cross-sectionally independent idiosyncratic noises.

An important class of spiked covariance models is factor models, which have been widely adopted in many disciplines, most notably, economics and finance. Some of the most famous examples include the capital asset pricing model (CAPM, Sharpe (1964)), the arbitrage pricing theory (Ross 1976), the approximate factor model (Chamberlain and Rothschild 1983), and the Fama-French three-factor model (Fama and French 1992).

The statistical analysis of both spiked covariance models and factor models fundamentally relies on Principal Component Analysis (PCA). One key inferential task, determining the number of factors, is equivalent to identifying the number of principal eigenvalues (Bai and Ng 2002; Onatski 2010; Ahn and Horenstein 2013). Furthermore, the estimation of factor loadings and the factors themselves depends on the principal eigenvectors (Stock and Watson 1998, 2002; Bai 2003; Bai and Ng 2006; Fan, Liao, and Mincheva 2011, 2013; Wang and Fan 2017).

In high-dimensional settings, Fan, Liao, and Mincheva (2013) establish a crucial theoretical link, demonstrating that factor analysis is asymptotically equivalent to PCA. Their work shows that under a "pervasiveness" condition, the space spanned by the principal components is asymptotically indistinguishable from the space spanned by the factor loadings; see Propositions 1 and 2 therein for the precise statements.

While factor models offer a parsimonious representation of high-dimensional dynamics, their estimation via PCA typically rests on a critical assumption: the time invariance of factor loadings. Inconsistent estimates arise if this assumption is violated. However, parameter instability is a pervasive feature of economic and financial time series, often driven by policy regime

CONTACT Xinghua Zheng xhzheng@ust.hk Department of Information System, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
#Additional affiliation: Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

shifts, technological shocks, or changes in economic fundamentals. The potential for such structural breaks necessitates careful diagnostic testing in empirical applications.

Statistically, a structural break manifests as a change in the factor loading matrix. While a substantial body of literature focuses on detecting the timing of such breaks (e.g., Stock and Watson 2009; Breitung and Eickmeier 2011; Andrew and Allan 2012; Chen, Dolado, and Gonzalo 2014; Han and Inoue 2015; Duan, Bai, and Han 2023; Bai, Duan, and Han 2025), less attention has been paid to diagnosing the *nature* of the break. The theoretical link provided in Fan, Liao, and Mincheva (2013) provides a powerful lens through which to analyze these changes. A structural break in the factor loadings can be due to a change in one or more principal eigenvalues, a change in one or more principal eigenvectors, or a change in both.

This distinction is of profound importance and provides a diagnostic signal about the nature of the instability. A change in a principal eigenvalue signals a shift in the absolute strength of its corresponding source of systematic variation. This, in the context of the stock market, indicates a significant change in the pricing of that specific source of risk. On the other hand, a change in a principal eigenvector points to a rotation in the orientation of the source of variation. In the context of the stock market, this signifies that the composition of the principal component, the economic or financial relationships it represents, has been redefined.

This critical distinction motivates the central objective of our paper: to develop a statistical framework for diagnosing the source of a suspected or detected structural break. Specifically, given two populations with covariance matrices $\mathbf{\Sigma}^{(1)}$ and $\mathbf{\Sigma}^{(2)}$, we propose tests for the following hypotheses:

(i) Test equality of principal eigenvalues: for each k, test

$$H_0^{(I,k)}: \lambda_k^{(1)} = \lambda_k^{(2)} \quad \text{versus} \quad H_a^{(I,k)}: \lambda_k^{(1)} \neq \lambda_k^{(2)};$$

(ii) Test equality of principal eigenvalue proportions: to account for potential heterogeneity in total variation across periods, test

$$H_0^{(II,k)}: \frac{\lambda_k^{(1)}}{\mathrm{tr}(\mathbf{\Sigma}^{(1)})} = \frac{\lambda_k^{(2)}}{\mathrm{tr}(\mathbf{\Sigma}^{(2)})} \quad \text{versus}$$
$$H_a^{(II,k)}: \frac{\lambda_k^{(1)}}{\mathrm{tr}(\mathbf{\Sigma}^{(1)})} \neq \frac{\lambda_k^{(2)}}{\mathrm{tr}(\mathbf{\Sigma}^{(2)})};$$

(iii) Test invariance of principal eigenvectors: test

$$H_0^{(III,k)}: |\langle \mathbf{v}_k^{(1)}, \mathbf{v}_k^{(2)} \rangle| = 1 \quad \text{vs} \quad H_a^{(III,k)}: |\langle \mathbf{v}_k^{(1)}, \mathbf{v}_k^{(2)} \rangle| < 1,$$

where $\langle \mathbf{a}, \mathbf{b} \rangle$ denotes the Euclidean inner product.

In the above, $\lambda_k^{(i)}$ and $\mathbf{v}_k^{(i)}$ stand for the kth largest eigenvalue and the corresponding eigenvector of $\mathbf{\Sigma}^{(i)}$, respectively. Among the three tests, the third one, concerning the stability of high-dimensional eigenvectors, is particularly challenging. This article makes a significant contribution by developing a rigorous test for this hypothesis.

Our work relates to, but is distinct from, several key streams of literature. The independent work of Bao et al. (2022) addresses similar questions but under a more restrictive setting. Our framework offers three significant generalizations. First, we allow for an arbitrary distribution of bulk eigenvalues, relaxing the assumption of equality imposed on bulk eigenvalues. Such a relaxation is important for practical applications. Second, our theory does not require the dimension-to-sample-size ratio to be bounded away from one. Third, we establish a two-sample Central Limit Theorem (CLT) for the inner product between two high-dimensional random eigenvectors, a more complex setting than their one-sample test of an eigenvector against a fixed direction.

Furthermore, our contribution differs fundamentally from the seminal works of Bai (2003) and Stock and Watson (2002) on factor analysis. Their analyses focus on the asymptotic distributions of estimated factors *at a fixed time* or factor loadings *for a fixed subject*, both of which are fixed and finite-dimensional objects. In contrast, we study the principal eigenvectors, whose dimension grows with the number of cross-sectional units. This high-dimensional nature renders conventional asymptotic theory inapplicable and necessitates the development of new theoretical tools.

To demonstrate the practical power and interpretive value of our theory, we work with an important problem in finance: analyzing structural breaks in financial markets. We apply our diagnostic framework to daily returns of S&P 500 constituent stocks. Our analysis reveals that not all structural breaks are alike, with different crises exhibiting diverse diagnostic profiles. For instance, the 2008 financial crisis and the 2020 COVID-19 pandemic stand out as comprehensive, systemic shocks, triggering simultaneous changes in factor strength, relative importance, and orientation. In contrast, the instability during periods like late 2011 and 2016 followed a different pattern, suggesting a repricing of risk through changes in factor strength and importance but without a significant realignment of the underlying market factors. This application underscores how the proposed methodology empowers researchers to move beyond detecting a break to understanding its underlying structural characteristics, thereby providing a more insightful view of market dynamics.

The rest of the article is organized as follows. Section 2 formalizes the spiked covariance framework. Section 3 presents the central limit theorems for principal eigenvalues, eigenvalue proportions, and eigenvectors. Section 4 details the construction of our three two-sample tests. Section 5 validates the methods through simulations. Sections 6 presents the results of our empirical application to S&P 500 data. Section 7 concludes. All proofs are relegated to the Appendix.

Notation: we use the following notation in addition to what has been introduced above. For a symmetric $N \times N$ matrix $\mathbf{A}$, its *empirical spectral distribution* (ESD) is defined as

$$F^{\mathbf{A}}(x) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}(\lambda_j(\mathbf{A}) \leq x), \quad x \in \mathbb{R},$$

where $\mathbf{1}(\cdot)$ is the indicator function. The limit of ESD as $N \to \infty$, if it exists, is referred to as the *limiting spectral distribution*, or LSD for short. For any vector $\mathbf{a}$, $\mathbf{a}[k]$ denotes its kth entry. We use "$\xrightarrow{\mathcal{D}}$" to denote weak convergence.

2. Setting and Assumptions

We assume that $(\mathbf{y}_t)_{t=1}^T$ is a sequence of i.i.d. $N-$dimensional random vectors with mean zero and covariance matrix $\mathbf{\Sigma}$. Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of $\mathbf{\Sigma}$ in descending order, and $\mathbf{v}_1, \ldots, \mathbf{v}_N$ be the corresponding eigenvectors. Write $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ and $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_N)$. The spectral decomposition of $\mathbf{\Sigma}$ is given by $\mathbf{\Sigma} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\mathsf{T}$.

We make the following assumptions.

Assumption A:

The eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_r > \lambda_{r+1} \geq \cdots \geq \lambda_N$ satisfy that

(A.i) for the principal part, one has $\lim_{N\to\infty} \lambda_k/N = \theta_k \in (0, \infty)$ for $1 \leq k \leq r$, where $r > 1$ is a fixed integer and θ_k's are distinct.

(A.ii) for the bulk part, there exists a $C_0 < \infty$ such that $\lambda_j \leq C_0$ for $j > r$, and the empirical distribution of $\{\lambda_{r+1}, \ldots, \lambda_N\}$ tends to a distribution H.

Remark 1. Assumption (A.i) corresponds to the setting of strong factors in the factor model literature. When the magnitudes of principal eigenvalues are smaller, say λ_i is of order N^α for some $\alpha \in (0, 1)$, the convergence of sample principal components still holds with a convergence rate depending on α. In this article, we focus on the case when $\alpha = 1$ and leave the study of the case when $\alpha \in (0, 1)$ for future work.

Assumption B: The observations $(\mathbf{y}_i)_{i=1}^T$ can be written as $\mathbf{y}_i = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{z}_i$, where $\mathbf{z}_i = (\mathbf{z}_i[1], \mathbf{z}_i[2], \ldots, \mathbf{z}_i[N])^\mathsf{T}$, $i = 1, 2, \ldots, T$, are i.i.d. random vectors, and $\mathbf{z}_i[\ell], \ell = 1, \ldots, N$, are independent random variables with zero mean, unit variance and satisfying $\sup_N \max_{1\leq\ell\leq N} \mathrm{E}(\mathbf{z}_1[\ell])^4 < \infty$.

Remark 2. Assumption B covers the multivariate normal case and coincides with the idea of PCA. Specifically, if $\mathbf{y}_i$ follows a multivariate normal distribution, then $\mathbf{z}_i = \mathbf{\Lambda}^{-1/2}\mathbf{V}^\mathsf{T}\mathbf{y}_i$ is an N-dimensional standard normal random vector and Assumption B naturally holds. On the other hand, under the orthogonal basis $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_N)$, the coordinates of $\mathbf{y}_i$ are $(\sqrt{\lambda_1}\mathbf{z}_i[1], \ldots, \sqrt{\lambda_N}\mathbf{z}_i[N])$. Assumption B says that the coordinate variables are independent with mean zero and variance $\lambda_i, i = 1, \ldots, N$.

Assumption C: The dimension N and sample size T are such that $\rho_N := N/T \to \rho \in (0, \infty)$ as $N \to \infty$.

3. One-Sample Asymptotics

Let $\widehat{\mathbf{\Sigma}}_N$ be the sample covariance matrix defined as

$$\widehat{\mathbf{\Sigma}}_N = \frac{1}{T}\sum_{t=1}^{\mathsf{T}} \mathbf{y}_t\mathbf{y}_t^\mathsf{T}.$$

Denote its eigenvalues by $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_N$, and let $\widehat{\mathbf{v}}_1, \ldots, \widehat{\mathbf{v}}_N$ be the corresponding eigenvectors.

Theorem 1. Under Assumptions A–C, as $N \to \infty$, the principal eigenvalues converge weakly to a multivariate normal distribution:

$$\sqrt{T}\begin{pmatrix} \widehat{\lambda}_1/\lambda_1 - 1 \\ \vdots \\ \widehat{\lambda}_r/\lambda_r - 1 \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, \mathbf{\Sigma}_\lambda), \tag{1}$$

where $\mathbf{\Sigma}_\lambda = \mathrm{diag}(\sigma_{\lambda_1}^2, \ldots, \sigma_{\lambda_r}^2)$ is a diagonal matrix with $\sigma_{\lambda_k}^2 = \mathrm{E}(\mathbf{z}_1[k])^4 - 1$.

Remark 3. The marginal convergence in (1) has been established in Wang and Fan (2017) under a sub-Gaussian assumption. We generalize their result to joint convergence and under a weaker moment assumption. See also Cai, Han, and Pan (2020) for a related result under a different setting.

Remark 4. By Theorem 3, the variance $\sigma_{\lambda_k}^2$ can be consistently estimated by

$$\widehat{\sigma_{\lambda_k}^2} = \frac{1}{T(\widehat{\lambda}_k)^2}\sum_{t=1}^{T}(\widehat{\mathbf{v}}_k^\mathsf{T}\mathbf{y}_t)^4 - 1,$$

hence, a feasible CLT is readily available.

Theorem 2. Under Assumptions A–C, for each $1 \leq k \leq r$, we have that as $N \to \infty$

$$\sqrt{\frac{T}{\widehat{\sigma_{-k}^2}}}\left(\frac{\widehat{\lambda}_k}{\mathrm{tr}(\widehat{\mathbf{\Sigma}}_N) - \widehat{\lambda}_k} - \frac{\lambda_k}{\mathrm{tr}(\mathbf{\Sigma}) - \lambda_k}\right) \xrightarrow{\mathcal{D}} N(0, 1), \tag{2}$$

where

$$\widehat{\sigma_{-k}^2} = \frac{\widehat{\lambda}_k^2}{\left(\mathrm{tr}(\widehat{\mathbf{\Sigma}}_N) - \widehat{\lambda}_k\right)^2}\left[\widehat{\sigma_{\lambda_k}^2} + \frac{\sum_{j=1,\ldots,r,j\neq k}\widehat{\lambda}_j^2\,\widehat{\sigma_{\lambda_j}^2}}{\left(\mathrm{tr}(\widehat{\mathbf{\Sigma}}_N) - \widehat{\lambda}_k\right)^2}\right].$$

Remark 5. Theorem 2 can be used to construct confidence intervals for the proportion $\varrho_k := \lambda_k/\mathrm{tr}(\mathbf{\Sigma})$. This follows easily from (2) and the fact that, if we write $\widetilde{\varrho}_k = \lambda_k/(\mathrm{tr}(\mathbf{\Sigma}) - \lambda_k)$, then $\varrho_k = \widetilde{\varrho}_k/(1 + \widetilde{\varrho}_k)$, which is a strictly increasing function of $\widetilde{\varrho}_k$.

Theorem 3. Under Assumptions A–C, for each $1 \leq k \leq r$, the principal sample eigenvector $\widehat{\mathbf{v}}_k$ satisfies that as $N \to \infty$,

$$T\left(1 - \langle\mathbf{v}_k, \widehat{\mathbf{v}}_k\rangle^2 - \frac{1}{T\widehat{\lambda}_k}\sum_{j=r+1}^{N}\frac{\widehat{\lambda}_j}{(1 - \widehat{\lambda}_j/\widehat{\lambda}_k)^2}\right) \xrightarrow{\mathcal{D}} \sum_{i=1,\ldots,r,i\neq k}\omega_{ki}\cdot Z_i^2,$$

where $\omega_{ki} = \theta_k\theta_i/(\theta_k - \theta_i)^2$, which can be consistently estimated by

$$\widehat{\omega_{ki}} = \frac{\widehat{\lambda}_k\widehat{\lambda}_i}{(\widehat{\lambda}_k - \widehat{\lambda}_i)^2},$$

and Z_i's are i.i.d. standard normal random variables.

Remark 6. Wang and Fan (2017) established the convergence rate for $|\langle v_k, \hat{v}_k\rangle|$ under a sub-Gaussian assumption. Theorem 3 extends their results by deriving the limiting distribution of $1-|\langle v_k, \hat{v}_k\rangle|^2$, which quantifies the asymptotic variability of eigenvector alignment. While Wang and Fan (2017) provide a foundation for consistency, our result is essential for statistical inference.

The proofs of Theorems 1–3 are given in the Appendix.

4. Two-Sample Tests

We now discuss how to conduct the three tests mentioned in the Introduction.

Suppose that we have two groups of observations of the same dimension N:

$$\mathbf{y}_1^{(1)},\dots,\mathbf{y}_{T_1}^{(1)}, \quad \text{and} \quad \mathbf{y}_1^{(2)},\dots,\mathbf{y}_{T_2}^{(2)},$$

which are drawn independently from two populations with mean zero and covariance matrices $\boldsymbol{\Sigma}^{(1)}$ and $\boldsymbol{\Sigma}^{(2)}$. We assume that Assumption B holds for each group of observations. Moreover, analogous to Assumption C, we have

$$\lim_{N\to\infty}\frac{N}{T_i}=\rho_i\in(0,\infty), \quad i=1,2.$$

Finally, analogous to Assumption A, with the spectral decompositions of $\boldsymbol{\Sigma}^{(i)}$ given by

$$\boldsymbol{\Sigma}^{(i)}=\mathbf{V}^{(i)}\boldsymbol{\Lambda}^{(i)}\mathbf{V}^{(i)\mathsf{T}}, \qquad i=1,2,$$

where $\boldsymbol{\Lambda}^{(i)}=\operatorname{diag}(\lambda_1^{(i)},\dots,\lambda_{r_i}^{(i)},\lambda_{r_i+1}^{(i)},\dots,\lambda_N^{(i)})$, and $\mathbf{V}^{(i)}=(\mathbf{v}_1^{(i)},\dots,\mathbf{v}_{r_i}^{(i)},\mathbf{v}_{r_i+1}^{(i)},\dots\mathbf{v}_N^{(i)})$, we assume that, for each population, there are r_i principal eigenvalues, which satisfy

$$\lim_{N\to\infty}\frac{\lambda_k^{(i)}}{N}=\theta_k^{(i)}\in(0,+\infty), \qquad \text{for } 1\le k\le r_i,\ i=1,2;$$

while the remaining eigenvalues $\lambda_j^{(i)}, r_i+1\le j\le N$, are uniformly bounded and have a limiting empirical distribution. Let us emphasize that we allow for $r_1\neq r_2$. Moreover, $\boldsymbol{\Sigma}^{(1)}$ and $\boldsymbol{\Sigma}^{(2)}$ can have different limiting spectral distributions. Let $r=\min\{r_1,r_2\}$.

Naturally, our tests will be based on the sample covariance matrices

$$\widehat{\boldsymbol{\Sigma}}_N^{(i)}=\frac{1}{T_i}\sum_{j=1}^{T_i}\mathbf{y}_j^{(i)}\mathbf{y}_j^{(i)\mathsf{T}}, \quad i=1,2.$$

Write their spectral decompositions as

$$\widehat{\boldsymbol{\Sigma}}_N^{(i)}=\widehat{\mathbf{V}}^{(i)}\widehat{\boldsymbol{\Lambda}}^{(i)}(\widehat{\mathbf{V}}^{(i)})^{\mathsf{T}},$$

where

$$\widehat{\boldsymbol{\Lambda}}^{(i)}=\operatorname{diag}(\widehat{\lambda}_1^{(i)},\dots,\widehat{\lambda}_N^{(i)}),\ \ \widehat{\mathbf{V}}^{(i)}=(\widehat{\mathbf{v}}_1^{(i)},\dots,\widehat{\mathbf{v}}_N^{(i)}).$$

4.1. Testing Equality of Principal Eigenvalues

To test $H_0^{(I,k)}:\ \lambda_k^{(1)}=\lambda_k^{(2)}$ for $k\le r$, we use the following test statistic

$$T_{E_k}=\sqrt{\frac{T_1T_2}{T_1\widehat{(\sigma_{\lambda_k}^2)}^{(2)}+T_2\widehat{(\sigma_{\lambda_k}^2)}^{(1)}}}\cdot\left(\frac{\widehat{\lambda}_k^{(1)}}{\widehat{\lambda}_k^{(2)}}-1\right),$$

where

$$\widehat{(\sigma_{\lambda_k}^2)}^{(i)}=\frac{1}{(\widehat{\lambda}_k^{(i)})^2\,T_i}\sum_{j=1}^{T_i}\left((\widehat{\mathbf{v}}_k^{(i)})^{\mathsf{T}}\mathbf{y}_j^{(i)}\right)^4-1, \quad i=1,2.$$

Theorem 4. For $k\le r$, under null hypothesis $H_0^{(I,k)}$ and Assumptions A–C, as $N\to\infty$, the proposed test statistic T_{E_k} converges weakly to the standard normal distribution.

Theorem 4 follows directly from Theorem 1 and the Delta method.

4.2. Testing Equality of Principal Eigenvalue Proportions

The null hypothesis

$$H_0^{(II,k)}:\ \frac{\lambda_k^{(1)}}{\operatorname{tr}(\boldsymbol{\Sigma}^{(1)})}=\frac{\lambda_k^{(2)}}{\operatorname{tr}(\boldsymbol{\Sigma}^{(2)})},$$

is equivalent to

$$H_0^{(II,k)'}:\ \frac{\lambda_k^{(1)}}{\operatorname{tr}(\boldsymbol{\Sigma}^{(1)})-\lambda_k^{(1)}}=\frac{\lambda_k^{(2)}}{\operatorname{tr}(\boldsymbol{\Sigma}^{(2)})-\lambda_k^{(2)}}.$$

Based on such an observation, we propose the following test statistic

$$T_{P_k}:=\frac{\sqrt{N}\left(\dfrac{\widehat{\lambda}_k^{(1)}}{\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_N^{(1)})-\widehat{\lambda}_k^{(1)}}-\dfrac{\widehat{\lambda}_k^{(2)}}{\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_N^{(2)})-\widehat{\lambda}_k^{(2)}}\right)}{\sqrt{\dfrac{N}{T_1}\widehat{\sigma_{-k}^2}^{(1)}+\dfrac{N}{T_2}\widehat{\sigma_{-k}^2}^{(2)}}},$$

where

$$\widehat{\sigma_{-k}^2}^{(i)}=\frac{(\widehat{\lambda}_k^{(i)})^2}{\left(\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_N^{(i)})-\widehat{\lambda}_k^{(i)}\right)^2}\times\left(\widehat{\sigma_{\lambda_k}^2}^{(i)}+\frac{\sum_{j=1,\dots,r_i,j\neq k}(\widehat{\lambda}_j^{(i)})^2\,\widehat{\sigma_{\lambda_j}^2}^{(i)}}{\left(\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_N^{(i)})-\widehat{\lambda}_k^{(i)}\right)^2}\right), \quad i=1,2.$$

Theorem 5. For $k\le r$, under the null hypothesis $H_0^{(II,k)}$ and Assumptions A–C, as $N\to\infty$, the test statistic T_{P_k} converges weakly to the standard normal distribution.

Theorem 5 is a straightforward consequence of Theorem 2.

4.3. Testing Invariance of Principal Eigenvectors

To test $H_0^{(III,k)}: |\langle \mathbf{v}_k^{(1)}, \mathbf{v}_k^{(2)}\rangle| = 1$ for $k \le r$, we use the following test statistic

$$T_{V_k} := 2N\left(1 - |\langle \widehat{\mathbf{v}}_k^{(1)}, \widehat{\mathbf{v}}_k^{(2)}\rangle|\right) - \frac{N^2}{T_1(N-r_1)\widehat{\lambda}_k^{(1)}} \sum_{j=r_1+1}^{N} \widehat{\lambda}_j^{(1)} - \frac{N^2}{T_2(N-r_2)\widehat{\lambda}_k^{(2)}} \sum_{j=r_2+1}^{N} \widehat{\lambda}_j^{(2)}. \tag{3}$$

Theorem 6. For $k \le r$, under null hypothesis $H_0^{(III,k)}$ and Assumptions A–C, suppose further that $\boldsymbol{\Xi}_{11}^* := \lim_{N\to\infty}\left(\langle \mathbf{v}_s^{(1)}, \mathbf{v}_t^{(2)}\rangle\right)_{s=1,\ldots,r_1; t=1,\ldots,r_2}$ exist. Then, as $N \to \infty$, the test statistic T_{V_k} converges weakly as follows:

$$T_{V_k} \xrightarrow{\mathcal{D}} \mathbf{q}_k^{\mathsf{T}} \begin{pmatrix} \mathbf{I}_{r_1-1} & -\boldsymbol{\Xi}_{11,-k}^* \\ (-\boldsymbol{\Xi}_{11,-k}^*)^{\mathsf{T}} & \mathbf{I}_{r_2-1} \end{pmatrix} \mathbf{q}_k,$$

where $\mathbf{q}_k$ is a $(r_1 + r_2 - 2)$-dimensional random vector, which follows a multivariate normal distribution with mean zero and covariance matrix

$$\mathbf{D}_k = \operatorname{diag}\Big(\rho_1\omega_{k1}^{(1)}, \ldots, \rho_1\omega_{k(k-1)}^{(1)}, \rho_1\omega_{k(k+1)}^{(1)}, \ldots, \rho_1\omega_{kr_1}^{(1)}, \rho_2\omega_{k1}^{(2)}, \ldots, \rho_2\omega_{k(k-1)}^{(2)}, \rho_2\omega_{k(k+1)}^{(2)}, \ldots, \rho_2\omega_{kr_2}^{(2)}\Big),$$

$$\omega_{kj}^{(i)} = \frac{\theta_k^{(i)}\theta_j^{(i)}}{(\theta_k^{(i)} - \theta_j^{(i)})^2}, \quad \text{for } j = 1, \ldots, k-1, k+1, \ldots, r_i,\ i = 1, 2,$$

and $\boldsymbol{\Xi}_{11,-k}^*$ is the matrix obtained by deleting the kth row and kth column of $\boldsymbol{\Xi}_{11}^*$. Furthermore, $\omega_{kj}^{(i)}$ and $\boldsymbol{\Xi}_{11}^*$ can be consistently estimated by

$$\widehat{\omega_{kj}^{(i)}} = \frac{\widehat{\lambda}_k^{(i)}\widehat{\lambda}_j^{(i)}}{(\widehat{\lambda}_k^{(i)} - \widehat{\lambda}_j^{(i)})^2}, \quad \text{and} \quad \widehat{\boldsymbol{\Xi}_{11}^*} = \left(\langle \widehat{\mathbf{v}_s^{(1)}}, \widehat{\mathbf{v}_t^{(2)}}\rangle\right)_{s=1,\ldots,r_1; t=1,\ldots,r_2},$$

respectively.

Corollary 1. Under the stronger null hypothesis that $r_1 = r_2$ and $|\langle \mathbf{v}_k^{(1)}, \mathbf{v}_k^{(2)}\rangle| = 1$ for all $k = 1, \ldots, r$, we have $\boldsymbol{\Xi}_{11}^* = \mathbf{I}_r$, and the test statistic T_{V_k} converges as follows:

$$T_{V_k} \xrightarrow{\mathcal{D}} \sum_{j=1,\ldots,r, j\neq k} \left(\rho_1\omega_{kj}^{(1)} + \rho_2\omega_{kj}^{(2)}\right) \cdot Z_j^2,$$

where Z_j's are i.i.d. standard normal random variables.

Theorem 6 and Corollary 1 are proved in the Appendix.

5. Simulation Studies

5.1. Design

We consider five covariance matrices as follows:

$$\boldsymbol{\Sigma}_1 = \mathbf{V}^{(1)}\boldsymbol{\Lambda}^{(1)}(\mathbf{V}^{(1)})^{\mathsf{T}}, \quad \boldsymbol{\Sigma}_2 = \mathbf{V}^{(1)}\boldsymbol{\Lambda}^{(2)}(\mathbf{V}^{(1)})^{\mathsf{T}},$$
$$\boldsymbol{\Sigma}_3 = \mathbf{V}^{(2)}\boldsymbol{\Lambda}^{(1)}(\mathbf{V}^{(2)})^{\mathsf{T}},$$
$$\boldsymbol{\Sigma}_4 = \boldsymbol{\Lambda}^{(1)}, \qquad \boldsymbol{\Sigma}_5 = \mathbf{V}^{(5)}\boldsymbol{\Lambda}^{(1)}(\mathbf{V}^{(5)})^{T},$$

where $\mathbf{V}^{(1)}, \mathbf{V}^{(2)}$ are two random orthogonal matrices, and

$$\boldsymbol{\Lambda}^{(1)} = \operatorname{diag}(2N, N, \lambda_3^{(1)}, \ldots, \lambda_N^{(1)}), \quad \lambda_j^{(1)} \sim_{\text{iid}} \text{Unif}(1, 3),$$
$$\boldsymbol{\Lambda}^{(2)} = \operatorname{diag}(7.5N, 1.5N, \lambda_3^{(2)}, \ldots, \lambda_N^{(2)}), \quad \lambda_j^{(2)} \sim_{\text{iid}} \text{Unif}(4.5, 7.5),$$
$$\mathbf{V}^{(5)} = (\mathbf{v}_1^{(5)}, \mathbf{v}_2^{(5)}, \mathbf{e}_3, \ldots, \mathbf{e}_N), \quad \text{where}$$
$$\mathbf{v}_1^{(5)} = (\cos\theta, \sin\theta, 0, \ldots, 0)^T,$$
$$\mathbf{v}_2^{(5)} = (-\sin\theta, \cos\theta, 0, \ldots, 0)^T, \quad \theta \in [0, \pi/2].$$

The observations will be simulated as follows: for a given $\boldsymbol{\Sigma}$, which will be one of the five covariance matrices above, write its spectral decomposition as $\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}}$. Then we simulate observations with covariance matrix $\boldsymbol{\Sigma}$ by $\mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{z}_t$, where $\mathbf{z}_t = (\mathbf{z}_t[1], \mathbf{z}_t[2], \ldots, \mathbf{z}_t[N])^{\mathsf{T}}$ consists of i.i.d. standardized student $t(8)$ random variables.

In Section 5.4, we report a set of results for the setting where observations are serially dependent. The results show that our tests work well in the presence of serial dependence and infinite kurtosis.

Theorems 4–6 are associated with different null hypotheses. When evaluating the sizes of the tests proposed in these theorems, we adopt the following setting:

- For both Theorems 4 and 5, we simulate two samples of observations with $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_3$ as their respective population covariance matrices. Note that $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_3$ share the same eigenvalues but have different eigenvectors.
- For Theorem 6, the two samples of observations are simulated with $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ as their respective population covariance matrices. The two matrices share the same eigenvectors but have different eigenvalues.

On the other hand, when evaluating powers, we use the following design:

- For testing equality of eigenvalues/eigenvalue proportions, the two samples of observations are simulated with $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ as their respective population covariance matrices;
- For testing invariance of principal eigenvectors, we simulate two samples of observations with $\boldsymbol{\Sigma}_4$ and $\boldsymbol{\Sigma}_5$ as their respective population covariance matrices. The difference between the principal eigenvectors of the two matrices is a function of the angle θ. We will change the value of θ to see how the power varies as a function of θ.

5.2. Visual Check

We first visually examine Theorems 4–6 by comparing the empirical distributions of the test statistics with their respective asymptotic distributions under the null hypotheses.

For Theorem 4, the asymptotic distribution of the test statistic T_{E_1} under the null hypothesis is the standard normal distribution. This is clearly supported by Figure 1, which gives the normal Q-Q plot and histogram of T_{E_1} based on 1,000 replications.

For Theorem 5, the asymptotic distribution of the test statistic T_{P_1} is again the standard normal distribution. This is supported by Figure 2.

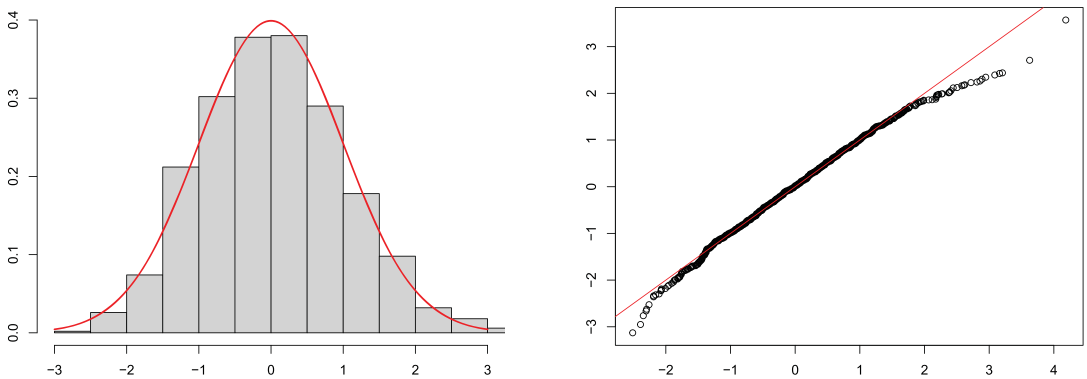

Figure 1. Histogram and Normal Q-Q plot of T_{E_1} based on 1,000 replications with $N = 500$, $T_1 = 500$ and $T_2 = 750$.

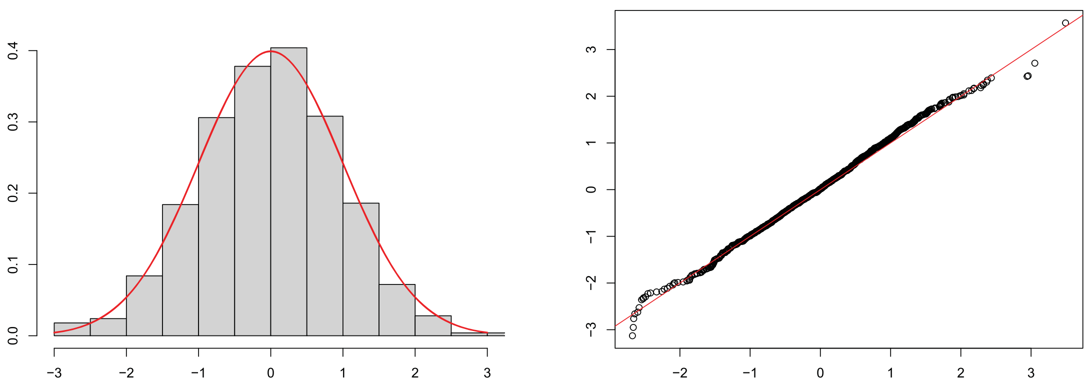

Figure 2. Histogram and Normal Q-Q plot of T_{P_1} based on 1000 replications with $N = 500$, $T_1 = 500$ and $T_2 = 750$.

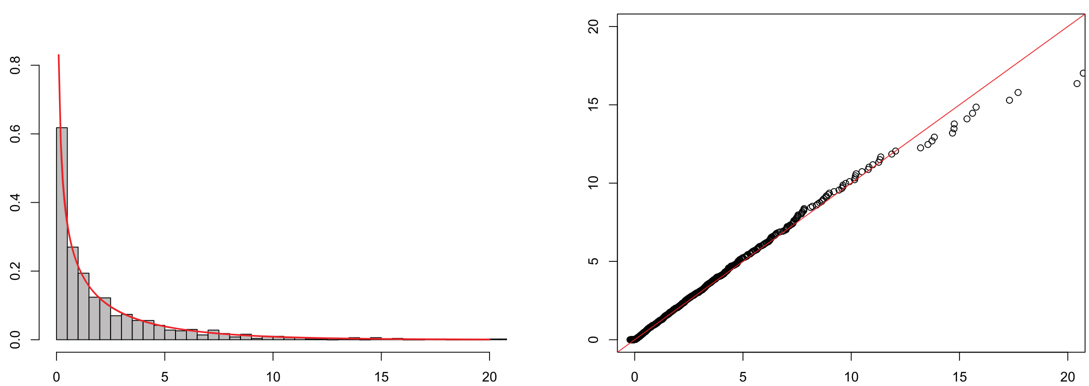

Figure 3. Comparisons of the empirical distribution of the test statistic T_{V_1} based on 1000 replications with the asymptotic distribution when $N = 500$, $T_1 = 500$, and $T_2 = 750$. Left: histogram of T_{V_1} versus the density of the asymptotic distribution; right: Q-Q plot of T_{V_1} versus the asymptotic distribution.

For Theorem 6, the test statistic T_{V_1} converges weakly to a quadratic form in normal random variables. The asymptotic distribution, in general, does not have an explicit density formula. Fortunately, Imhof (1961) provides precise numerical methods to compute the density and quantiles.[1] In Figure 3, we compare the empirical distribution of the test statistic T_{V_1} with the asymptotic distribution. We see that the histogram matches well the density of the asymptotic distribution, and the Q-Q plot exhibits a good agreement.

5.3. Size and Power Evaluation

In this subsection, we evaluate the sizes and powers of the three tests presented in Theorems 4–6. The significance level is set to be 5%. Tests based on T_{P_k} and T_{V_k} involve the number of factors,

[1] We thank Raymond Kan for bringing the paper to our attention and for sharing the code to compute the density and quantiles.

which are unknown in practice. There are several estimators available, including those given in Bai and Ng (2002), Onatski (2010), and Ahn and Horenstein (2013). The results below are based on the estimator in Bai and Ng (2002). For tests based on T_{E_k} and T_{P_k}, the critical values are standard normal quantiles. For test based on T_{V_k}, the critical value is the 95% quantile of the asymptotic distribution, which we compute using the method in Imhof (1961).

Table 1 reports the empirical sizes of the three tests based on T_{E_k}, T_{P_k} and T_{V_k}, $k = 1, 2,$, for different combinations of N, T_1 and T_2. We see that for the first two sets of tests, for different N and T_1, T_2, the empirical sizes are close to the nominal level of 5%. For the third set of tests based on T_{V_k}, $k = 1, 2$, the size approaches 5% as the dimension N and sample sizes T_1, T_2 get larger.

Power evaluation results are given in Table 2. We see that the powers reach almost one as the dimension N increases to 500.

Finally, in Figure 4, we evaluate the power of the eigenvector test $T_{V_k}, k = 1, 2$ as a function of θ. For the three θ values tested, $i\pi/9, i = 1, 2, 3$, the bigger the value, the bigger the difference between the principal eigenvectors in the two populations, and the higher the power. Moreover, even for the smallest value $\pi/9$, the power quickly increases to close to 1 as the dimension N and sample sizes T_1, T_2 get larger.

Table 1. Empirical sizes of T_{E_k}, T_{P_k}, and T_{V_k}, $k = 1, 2$, based on 1000 replications at 5% significance level with $N/T_1 = 1$ and $N/T_2 = 2/3$.

N	T_{E_1}	T_{E_2}	T_{P_1}	T_{P_2}	T_{V_1}	T_{V_2}
100	0.064	0.063	0.043	0.049	0.092	0.087
300	0.037	0.050	0.037	0.037	0.063	0.060
500	0.062	0.053	0.052	0.047	0.061	0.065

Table 2. Empirical powers of T_{E_k}, T_{P_k}, and T_{V_k} (for $\theta = \pi/9$), $k = 1, 2$, based on 1000 replications at 5% significance level with $N/T_1 = 1$ and $N/T_2 = 2/3$.

N	T_{E_1}	T_{E_2}	T_{P_1}	T_{P_2}	T_{V_1}	T_{V_2}
100	0.977	0.319	0.433	0.789	0.480	0.485
300	0.999	0.768	0.848	0.998	0.888	0.888
500	1.000	0.939	0.951	1.000	0.986	0.987

5.4. Simulation Results under a Serially Dependent Setting

In this subsection, we evaluate the applicability of our proposed methods to time series data with serial dependence.

We assume that the observations are generated by

$$\begin{aligned} \mathbf{y}_t^{(i)} &= \widetilde{\mathbf{V}}^{(i)}(\widetilde{\boldsymbol{\Lambda}}^{(i)})^{1/2}\mathbf{z}_t^{(i)}, \qquad \text{where} \\ \mathbf{z}_t^{(i)} &= 0.1\, \mathbf{z}_{t-1}^{(i)} + \sqrt{1-(0.1)^2}\, \boldsymbol{\varepsilon}_t^{(i)}, \qquad i = 1, 2, \text{ and} \\ \boldsymbol{\varepsilon}_t &= (\boldsymbol{\varepsilon}_t[1], \boldsymbol{\varepsilon}_t[2], \ldots, \boldsymbol{\varepsilon}_t[N])^\top \end{aligned} \tag{4}$$

consists of i.i.d. standardized student $t(5)$ random variables.

Following the simulation design in Section 5.1, for the null hypothesis $H_0^{(I,k)}$ and $H_0^{(II,k)}$, we set

$$\widetilde{\mathbf{V}}^{(1)} = \mathbf{V}^{(1)}, \quad \widetilde{\mathbf{V}}^{(2)} = \mathbf{V}^{(2)}, \quad \widetilde{\boldsymbol{\Lambda}}^{(1)} = \widetilde{\boldsymbol{\Lambda}}^{(2)} = \boldsymbol{\Lambda}^{(1)}.$$

For the null hypothesis $H_0^{(III,k)}$, we set

$$\widetilde{\mathbf{V}}^{(1)} = \widetilde{\mathbf{V}}^{(2)} = \mathbf{V}^{(1)}, \quad \widetilde{\boldsymbol{\Lambda}}^{(1)} = \boldsymbol{\Lambda}^{(1)}, \quad \widetilde{\boldsymbol{\Lambda}}^{(2)} = \boldsymbol{\Lambda}^{(2)}.$$

For the alternative hypothesis $H_1^{(I,k)}$ and $H_1^{(II,k)}$, we set

$$\widetilde{\mathbf{V}}^{(1)} = \widetilde{\mathbf{V}}^{(2)} = \mathbf{V}^{(1)}, \quad \widetilde{\boldsymbol{\Lambda}}^{(1)} = \boldsymbol{\Lambda}^{(1)}, \quad \widetilde{\boldsymbol{\Lambda}}^{(2)} = \boldsymbol{\Lambda}^{(2)}.$$

For the alternative hypothesis $H_1^{(III,k)}$, we set

$$\widetilde{\mathbf{V}}^{(1)} = \mathbf{I}_N, \quad \widetilde{\mathbf{V}}^{(2)} = \mathbf{V}^{(5)}, \quad \widetilde{\boldsymbol{\Lambda}}^{(1)} = \widetilde{\boldsymbol{\Lambda}}^{(2)} = \boldsymbol{\Lambda}^{(1)}.$$

Figures 5 and 6 present the histograms and Q-Q plots of T_{E_1} and T_{P_1} based on 1000 replications with $N = 500, T_1 = 500$ and $T_2 = 750$ under the null hypotheses. The histograms closely align with the standard normal density, and the Q-Q plots show a strong linear pattern. We see the distributions of T_{E_1} and T_{P_1} are still approximately standard normal even if the assumptions of Theorems 4 and 5 are violated.

Figure 7 displays the histogram and Q-Q plot of T_{V_1} against the asymptotic distribution under the null hypothesis. We see that the empirical distribution of the test statistic T_{V_1} still matches the asymptotic distribution quite well.

Table 3 reports the empirical sizes of T_E, T_P, and T_V at the 5% significance level for varying dimensions N. We see that the size improves as N gets larger and are roughly 5% when $N = 500$.

Table 4 summarizes the empirical powers of T_E, T_P, and T_V. We see once again that power increases with the dimension N and is quite high when $N = 500$.

Overall, these results suggest that our tests work well in the presence of both serial dependence and heavy-tailedness.

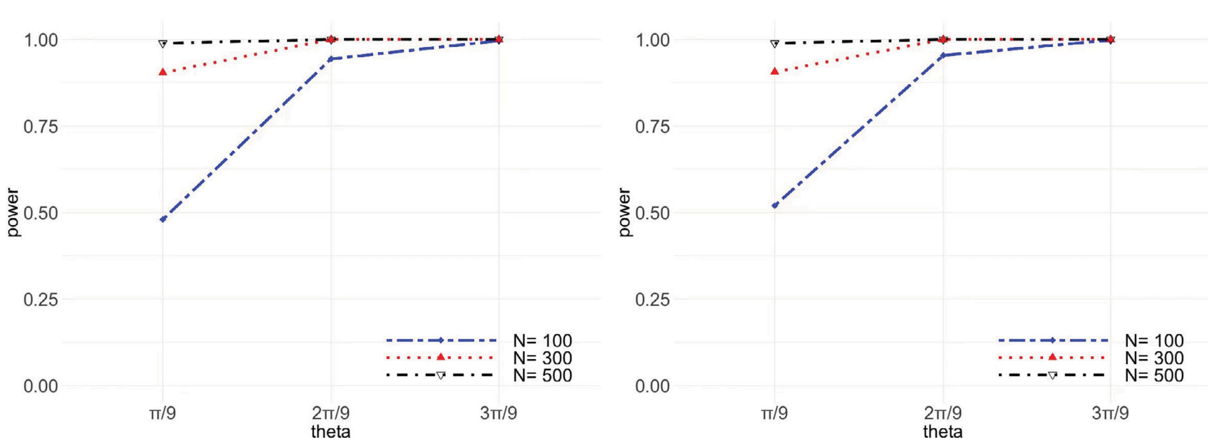

Figure 4. Empirical power of T_{V_i}, $i = 1, 2$, as a function of θ at 5% significance level for different N and T_1, T_2 with $N/T_1 = 1$ and $N/T_2 = 2/3$. Left: powers for T_{V_1}; right: powers for T_{V_2}.

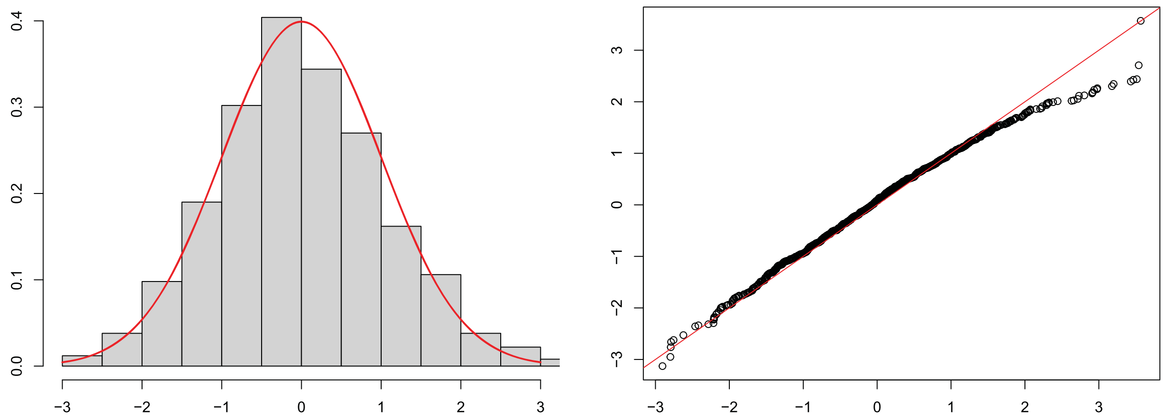

Figure 5. Histogram (left) and Normal Q-Q plot (right) of T_{E_1} under Model (4) under the null hypothesis based on 1000 replications with $N = 500$, $T_1 = 500$, and $T_2 = 750$.

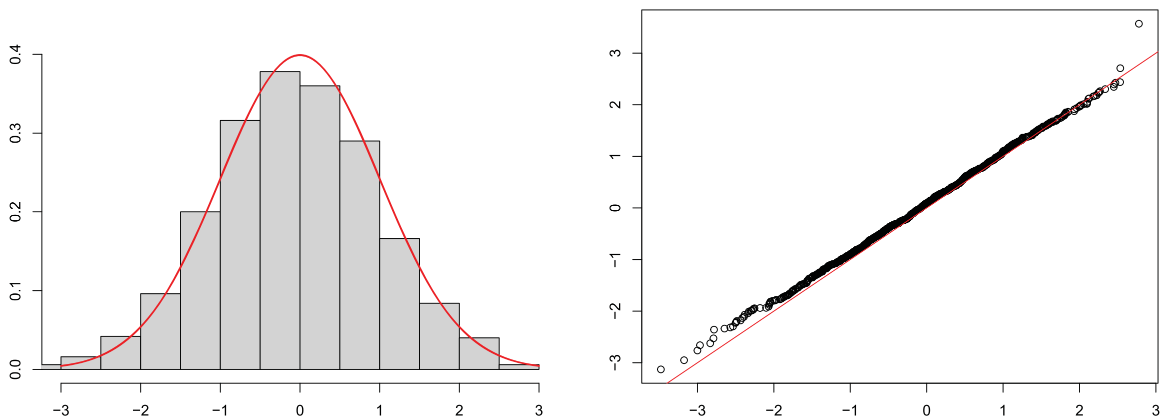

Figure 6. Histogram (left) and Normal Q-Q plot (right) of T_{P_1} under Model (4) under the null hypothesis based on 1000 replications with $N = 500$, $T_1 = 500$, and $T_2 = 750$.

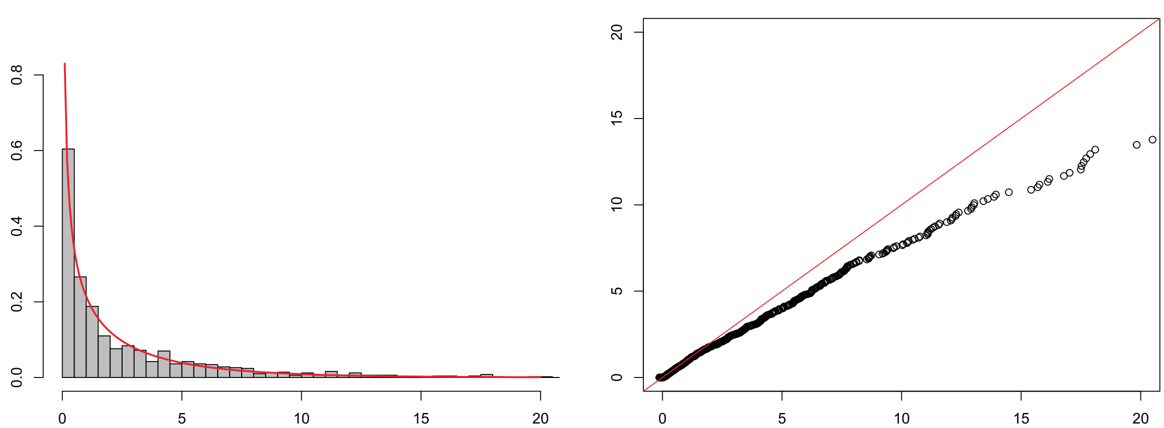

Figure 7. Comparisons of the empirical distribution of the test statistic T_{V_1} based on 1000 replications under Model (4) under the null hypothesis with the asymptotic distribution when $N = 500$, $T_1 = 500$, and $T_2 = 750$. Left: histogram of T_{V_1} versus the density of the asymptotic distribution; right: Q-Q plot of T_{V_1} versus the asymptotic distribution.

Table 3. Empirical sizes of T_{E_k}, T_{P_k}, and T_{V_k} under Model (4) based on 1000 replications at 5% significance level with $N/T_1 = 1$ and $N/T_2 = 2/3$.

N	T_{E_1}	T_{E_2}	T_{P_1}	T_{P_2}	T_{V_1}	T_{V_2}
100	0.069	0.076	0.034	0.061	0.108	0.111
300	0.072	0.057	0.054	0.049	0.080	0.086
500	0.062	0.071	0.052	0.058	0.068	0.069

Table 4. Empirical powers of T_{E_k}, T_{P_k}, and T_{V_k} under Model (4) based on 1000 replications for $\theta = \pi/9$ at 5% significance level with $N/T_1 = 1$ and $N/T_2 = 2/3$.

N	T_{E_1}	T_{E_2}	T_{P_1}	T_{P_2}	T_{V_1}	T_{V_2}
100	0.892	0.197	0.334	0.603	0.494	0.494
300	0.985	0.571	0.659	0.941	0.907	0.904
500	0.997	0.778	0.801	0.989	0.981	0.981

6. Empirical Studies

In this section, we apply our proposed two-sample tests to stock market data to demonstrate the effectiveness of our tests in diagnosing the nature of structural breaks. We first identify major market shifts based on existing structural break detection methods. We then apply our tests to determine whether the structural breaks stem from changes in factor strength, proportional variance, or fundamental orientations.

6.1. Data and Methodology

Our dataset comprises daily log-returns of constituent stocks from the S&P 500 Index, spanning from January 2007 to Dec 2024, which are obtained from the Center for Research in Security Prices (CRSP). We randomly select 200 stocks that remain in the S&P 500 constituents across the entire period.

To identify potential structural breaks, we use the quasi-maximum likelihood estimator of Bai, Duan, and Han (2025), which can detect multiple breakpoints.[2]

For each breakpoint detected, we construct the pre- and post-break data samples. To maximize sample integrity by avoiding data contamination from adjacent breaks while also ensuring the tests have sufficient statistical power, we adhere to the following rules:

- If the period to the nearest breakpoint is greater than one year on both sides, we use one year of data for each sample.
- If the period on one side is less than eight months, no test is performed for that breakpoint, as the sample size is insufficient for a reliable test.
- If one side has over a year of data while the other has between 8 and 12 months, we use one year of data for the longer side and the entire eligible data window (i.e., all data up to the adjacent breakpoint) for the shorter side. This adaptive window approach ensures we use all available "clean" data without crossing into a new market regime.

Finally, we conduct our two-sample tests using these pre- and post-break samples. Our tests involve the number of factors for each sample, which we estimate using the information criteria of Bai and Ng (2002). In the next subsection, we report the results for the first three principal components, with more discussion on the results for the first principal component (PC1) to clearly illustrate the framework's diagnostic capabilities.

6.2. Empirical Results and Discussions

We apply our three tests for eigenvalue equality (TE), eigenvalue proportion stability (TP), and eigenvector invariance (TV) to samples around the detected breakpoints to illustrate the diagnostic capabilities of the proposed framework. All tests are conducted at the 5% significance level. Same as in the Simulation Studies, for tests based on T_{E_k} and T_{P_k}, the critical values are standard normal quantiles; for test based on T_{V_k}, the critical value is the 95% quantile of the asymptotic distribution under the null hypothesis.

The results for the first three principal components, based on our sample of 200 stocks, are detailed in Table 5. Table 5 provides a granular diagnosis of the detected market shifts and reveals the distinct information captured by each test. Our tests suggest that changes in absolute factor strength (TE rejection) and factor orientation (TV rejection) are frequent characteristics of breaks affecting the first principal component (PC1). The tests also reveal that shifts in relative importance (TP rejection) are largely concentrated in PC1, while the relative contributions of PC2 and PC3 appear more stable. This highlights how the framework can disentangle the complex nature of structural changes.

Table 5. Test results for detected structural breakpoints.

Breakpoint	PC1			PC2			PC3		
	TE	TP	TV	TE	TP	TV	TE	TP	TV
2008-09-12	✗	✗	✗	✗	✓	✗	✓	✓	✗
2009-06-22	✗	✓	✗	✓	✓	✗	✓	✓	✗
2011-12-21	✗	✗	✓	✓	✓	✗	✓	✓	✗
2014-10-07	✓	✓	✗	✓	✓	✗	✓	✓	✗
2016-11-07	✗	✗	✓	✓	✓	✗	✓	✓	✗
2020-02-21	✗	✗	✗	✗	✓	✗	✓	✓	✓
2022-10-17	✗	✓	✗	✗	✓	✗	✗	✓	✓

NOTE: "PC Index" refers to the principal component being tested. "✗" indicates a rejection of the null hypothesis for eigenvalue equality (TE), proportion stability (TP), or eigenvector invariance (TV), and "✓" indicates retention. All tests are conducted at the 5% significance level.

These patterns become clearer when we focus on the first principal component (PC1), which typically represents the overall market factor, as visualized in Figure 8. The figure plots the S&P 500 Index alongside markers indicating the nature of each structural break for PC1. For the breaks corresponding to the 2008 financial crisis and the 2020 COVID-19 pandemic, our results show a simultaneous rejection of all three null hypotheses (TE, TP, and TV) for PC1 (●, ▲, ■). This specific and consistent combination of outcomes points to a potential dislocation across factor strength, relative importance, and orientation, illustrating how the framework can flag events that may be of a particularly comprehensive and systematic nature.

The framework's ability to provide a more nuanced diagnosis is illustrated by its application to other breakpoints. For example, the tests reveal a similar pattern for the breaks in late 2011 and 2016. In these cases, the results suggest that changes in factor strength and relative importance (TE and TP rejections ●, ▲) occurred without a significant change in factor orientation (TV retention). This pattern indicates a potential repricing of risk that is structurally different from a fundamental realignment of market factors. In another illustration, the break in 2014 is associated with a rejection for eigenvector invariance alone (■), suggesting that that period of instability may have been driven primarily by a shift in factor orientation.

7. Conclusions

This study establishes one-sample and two-sample central limit theorems for principal eigenvalues and eigenvectors under divergent spiked covariance models, laying a rigorous theoretical foundation for diagnosing structural breaks in

[2]We thank Xu Han for sharing the code with us.

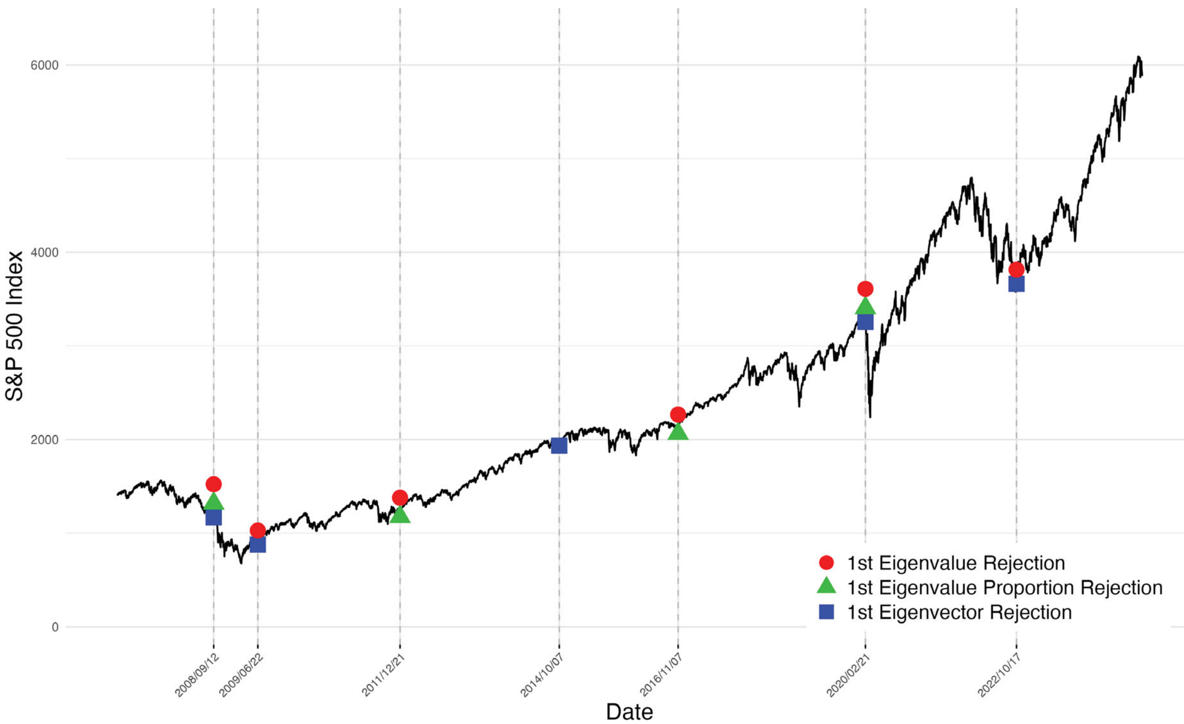

Figure 8. Test results regarding the 1st principal component. Vertical lines mark the dates of detected structural breaks. Symbols indicate rejections of the null hypotheses for the first principal component (PC1): a red circle (●) for eigenvalue equality (T_E), a green triangle (▲) for eigenvalue ratio consistency (T_P), and a blue square (■) for eigenvector invariance (T_V). All tests are conducted at the 5% significance level.

high-dimensional data. The core contribution lies in the development of three two-sample tests that collectively allow researchers to investigate the source of a structural break, distinguishing whether it arises from changes in factor strength (eigenvalues), relative factor importance (eigenvalue proportions), or fundamental factor orientation (eigenvectors).

Numerical validation through simulation studies confirms the robustness of the proposed tests in finite samples. Our application to daily returns of S&P 500 constituent stocks illustrates the framework's diagnostic utility. It shows how the tests can identify periods of major systematic change, characterized by simultaneous shifts across all three structural aspects, and distinguish them from more contained structural breaks. Furthermore, the application demonstrates how the framework can identify more nuanced patterns of instability, such as periods where tests suggest changes in factor strength and importance but not orientation, versus others that appear to be driven primarily by a realignment of factors alone.

This ability to provide a differentiated diagnosis of instability provides a powerful tool for structural break analysis, bridging theoretical advances in high-dimensional statistics with actionable insights for empirical research.

Supplementary Materials

The supplementary materials include Appendix, which contains proofs and code files.

Acknowledgments

We are grateful to the Editor, Associate Editor and two anonymous referees for their thoughtful and constructive suggestions, which led to significant improvements of the article. We also thank Haokun Yang and Ruizhao Huang for excellent research assistance.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Research is partially supported by HKSAR RGC grants SRFS2324-6H02, 16304521, and NSFC(No. 72531004).

Data Availability Statement

The data in the Empirical Studies are obtained from the Center for Research in Security Prices (CRSP).

References

Ahn, S. C., and Horenstein, A. R. (2013), "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227. [1,7]

Andrew, A., and Allan, T. (2012), "Regime Changes and Financial Markets," *Annual Review of Financial Economics*, 4, 313–337. [2]

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171. [1,2]

Bai, J., Duan, J., and Han, X. (2025), "Singularity-Based Consistent QML Estimation of Multiple Breakpoints in High-Dimensional Factor Models," working paper. [2,9]

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [1,7,9]

——— (2006), "Evaluating Latent and Observed Factors in Macroeconomics and Finance," *Journal of Econometrics*, 131, 507–537. [1]

Bao, Z., Ding, X., Wang, J., and Wang, K. (2022), "Statistical Inference for Principal Components of Spiked Covariance Matrices," *The Annals of Statistics*, 50, 1144–1169. [1,2]

Breitung, J., and Eickmeier, S. (2011), "Testing for Structural Breaks in Dynamic Factor Models," *Journal of Econometrics*, 163, 71–84. [2]

Cai, T. T., Han, X., and Pan, G. (2020), "Limiting Laws for Divergent Spiked Eigenvalues and Largest Nonspiked Eigenvalue of Sample Covariance Matrices," *The Annals of Statistics*, 48, 1255–1280. [1,3]

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304. [1]

Chen, L., Dolado, J. J., and Gonzalo, J. (2014), "Detecting Big Structural Breaks in Large Factor Models," *Journal of Econometrics*, 180, 30–48. [2]

Duan, J., Bai, J., and Han, X. (2023), "Quasi-Maximum Likelihood Estimation of Break Point in High-Dimensional Factor Models," *Journal of Econometrics*, 233, 209–236. [2]

Fama, E. F., and French, K. R. (1992), "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47, 427–465. [1]

Fan, J., Liao, Y., and Mincheva, M. (2011), "High-Dimensional Covariance Matrix Estimation in Approximate Factor Models," *The Annals of Statistics*, 39, 3320–3356. [1]

——— (2013), "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society*, Series B, 75, 603–680, with 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva. [1,2]

Fan, J., Wang, W., and Zhu, Z. (2021), "A Shrinkage Principle for Heavy-Tailed Data: High-Dimensional Robust Low-Rank Matrix Recovery," *Annals of statistics*, 49, 1239–1266. [1]

Fan, J., Yan, Y., and Zheng, Y. (2024), "When Can Weak Latent Factors be Statistically Inferred?" arXiv preprint arXiv:2407.03616. [1]

Han, X., and Inoue, A. (2015), "Tests for Parameter Instability in Dynamic Factor Models," *Econometric Theory*, 31, 1117–1152. [2]

Imhof, J. P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426. [6,7]

Johnstone, I. M. (2001), "On the Distribution of the Largest Eigenvalue in Principal Components Analysis," *The Annals of Statistics*, 29, 295–327. [1]

Li, H., Aue, A., Paul, D., and Peng, J. (2024), "Testing General Linear Hypotheses under a High-Dimensional Multivariate Regression Model with Spiked Noise Covariance," *Journal of the American Statistical Association*, 119, 2799–2810. [1]

Onatski, A. (2010), "Determining the Number of Factors from Empirical Distribution of Eigenvalues," *The Review of Economics and Statistics*, 92, 1004–1016. [1,7]

Ross, S. (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360. [1]

Sharpe, W. (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425–442. [1]

Stock, J. H., and Watson, M. W. (1998), "Diffusion Indexes," Working Paper, 6702. [1]

——— (2002), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179. [1,2]

——— (2009), "Forecasting to Dynamic Factor Models Subject to Structural Instability," in *The Methodology and Practice of Econometrics*, pp. 173–205, Oxford: Oxford University Press. [2]

Wang, B., and Fan, J. (2025), "Robust Matrix Completion with Heavy-Tailed Noise," *Journal of the American Statistical Association*, 120, 922–934. [1]

Wang, W., and Fan, J. (2017), "Asymptotics of Empirical Eigenstructure for High Dimensional Spiked Covariance," *The Annals of Statistics*, 45, 1342–1374. [1,3,4]

Yan, Y., Chen, Y., and Fan, J. (2024), "Inference for Heteroskedastic PCA with Missing Data," *The Annals of Statistics*, 52, 729–756. [1]

Zhang, A. R., Cai, T. T., and Wu, Y. (2022), "Heteroskedastic PCA: Algorithm, Optimality, and Applications," *The Annals of Statistics*, 50, 53–80. [1]