Department of Econometrics and Business Statistics

<http://business.monash.edu/econometrics-and-business-statistics/research/publications>

Spiked Eigenvalues of High-Dimensional Separable Sample Covariance Matrices

Bo Zhang, Jiti Gao, Guangming Pan and
Yanrong Yang

December 2019

Working Paper 31/19

Spiked Eigenvalues of High-Dimensional Separable Sample Covariance Matrices*

Bo Zhang¹, Jiti Gao², Guangming Pan³, Yanrong Yang⁴

¹University of Science and Technology of China, China

²Monash University, Australia

³Nanyang Technological University, Singapore

⁴The Australian National University, Australia

December 2, 2019

Abstract

This paper establishes asymptotic properties for spiked empirical eigenvalues of sample covariance matrices for high-dimensional data with both cross-sectional dependence and a dependent sample structure. A new finding from the established theoretical results is that spiked empirical eigenvalues will reflect the dependent sample structure instead of the cross-sectional structure under some scenarios, which indicates that principal component analysis (PCA) may provide inaccurate inference for cross-sectional structures. An illustrated example is provided to show that some commonly used statistics based on spiked empirical eigenvalues misestimate the true number of common factors. As an application of high-dimensional time series, we propose a test statistic to distinguish the unit root from the factor structure and demonstrate its effective finite sample performance on simulated data. Our results are then applied to analyze OECD healthcare expenditure data and U.S. mortality data, both of which possess cross-sectional dependence as well as non-stationary temporal dependence. It is worth mentioning that we contribute to statistical justification for the benchmark paper by Lee and Carter [25] in mortality forecasting.

Keywords: Factor model; high-dimensional data; principal component analysis; spiked empirical eigenvalue.

JEL Classification: C21, C32, C55.

*The research of the first and the second authors was supported by the Australian Research Council Discovery Grants Program under Grant Number: DP170104421.

1 Introduction

With the rapid development of computer science, data from various scientific areas demonstrate the following features: high dimensionality, cross-sectional dependence, and a dependent sample structure (including serial dependence in the time series setting). Both high-dimensional time series and high-dimensional spatial data possess these features, which are commonly encountered in modern statistical analysis. Taking human mortality data from actuarial science and demography as an example (see Lee and Carter [25]), age-specific human mortality data is a high-dimensional time series, in which death rates for each age are observed for several years and the number of ages (i.e. the dimension) is comparable to the length of time observations (i.e. the sample size). Due to the strong relationship between the death rates at different ages, cross-sectional dependence exists in this data. Meanwhile, mortality data indicate non-stationary temporal trending behavior. Another popular dataset is healthcare expenditure data accumulated from certain countries and observed over several years, which are usually modeled as panel data in econometrics (see Gao, Xia and Zhu [18]). This dataset illustrates strong cross-sectional dependence across countries, as well as a non-stationary tendency in the temporal direction. Complicated structures among sample observations incur a new curse for high-dimensional statistical analysis, in parallel with the curse of dimensionality.

As is well known, the curse of dimensionality has been dealt with extensively in the high-dimensional statistical literature over the last few decades. Dimensional reduction and variable selection are the major techniques for reducing the exploding flexibility in traditional statistical approaches for high-dimensional data. While most of these statistical research works focus on independent and identically distributed (i.i.d) sampled data, they are rarely extended to high-dimensional data with dependent sample observations, especially non-stationary time series. We will contribute to this area by analyzing the influence of dependent sample observations on PCA for high-dimensional data. The interplay between high dimensionality and dependent sample structure may lead to new behavior of the spiked empirical eigenvalues and limit the feasibility of PCA. PCA is a traditional approach for reducing the dimensionality of data by representing it in terms of new variables, called principal components, that largely capture some meaningful variation in the data. An early effort on the establishment of PCA theory is from Anderson [1]. From a technical point of view, PCA intends to use spiked empirical eigenstructures (eigenvalues and eigenvectors) from sample covariance matrices to estimate or infer the spiked cross-sectional population eigenstructures (cross-sectional dependence). Here, the spiked empirical eigenvalues refer to those leading sample eigenvalues well separated from the remaining sample eigenvalues, which were first termed as such in Johnstone [22].

Let us look at an example to show that spiked empirical eigenvalues play an important role in PCA. Factor model is equivalent to PCA to some extent; see for example, Fan, Liao, and Mincheva [15]. Estimation of the number of factors in a factor model is important for statistical inference. The existing literature provides some useful methods driven by spiked empirical eigenvalues. Bai [3] estimates the number of factors based on the largest empirical eigenvalues and penalty functions. Briefly, the idea is that the number of factors is the number of the empirical eigenvalues that are larger than the penalty function. Anh and Horensten [2] estimate the number of factors from a

different point of view: if the number of factors is k , the ratio between the k -th largest empirical eigenvalue and the $(k + 1)$ -th largest empirical eigenvalue would be larger than others. Onatski [29] considers another idea: if the number of factors is k , for any $s > k$, the difference between the s -th largest empirical eigenvalue and the $(s + 1)$ -th one should be very small. These papers work well in factor models with their assumptions. If we apply these methods to certain general cases, it is necessary to investigate spiked eigenvalues from sample covariance matrices under certain conditions.

Random matrix theory is an important tool for studying high-dimensional sample covariance matrices and has provided many important theoretical results on empirical eigenvalues and the corresponding empirical eigenvectors. Many researchers have further developed useful estimation and testing procedures based on these empirical spectral results on high-dimensional data. Recent discussions in this area can be found in Bai and Silverstein[4], Johnstone [23], Paul and Aue [32] and Yao, Zheng, and Bai [38]. Research into empirical eigenvalues dates back to 1930s (see Fisher [17], Hsu [21] and Roy [33]). The earliest researchers focused on the fixed dimensional case. It is well known that sample covariance matrices behave significantly differently to those for the fixed dimensional case as the dimension goes to infinity. For instance, Marcenko and Pastur [27] establish a limiting spectral distribution (MP law) for the sample covariance matrix as the dimension (n) and sample size (T) are comparable. Johnstone [22] shows that the largest eigenvalue of the sample covariance matrices follows the Tracy-Widom law, which is quite different from the results in the low-dimensional case. Since then, the asymptotic properties of the eigenvalues of high-dimensional sample covariance matrices have attracted many researchers to contribute in this area. Among them, we mention Baik et al.[6], Baik and Silverstein [7], Paul [31], Bai and Yao [5], and Cai, Han, and Pan [10], who have studied spiked empirical eigenvalues. The results on spiked empirical eigenvalues in the literature reveal one-to-one correspondence between the spiked empirical eigenvalues and the spiked population eigenvalues, which guarantees the feasibility of PCA on high-dimensional data with independent samples.

Most of the existing studies rely on the assumption that the observations of high-dimensional data are independent, although cross-sectional dependence can be allowed. It is noted that when high-dimensional data appear to be nonstationary, the first differenced version of the data is normally used for statistical estimation and inference. However, observations of high-dimensional data in economics and finance, for example, are often temporally dependent and even highly nonstationary, and they don't necessarily follow a standard type of unit-root nonstationarity. Our experience shows that differencing may not be a good idea even though it is convenient. First, differenced versions may not necessarily be stationary. Second, differencing original data may simply ignore some key features of the original data, such as trending behaviours, which are often the main interest in analysing some economic and financial time series. In the end of Section 5.3, we make some specific comments on this. Zhang, Pan, and Gao [39] were among the first dealing with the largest eigenvalues of sample covariance matrices generated from high-dimensional nonstationary time series. They find that dependent sample observations are also able to result in spiked empirical eigenvalues even with the absence of cross-sectional dependence. This phenomenon indicates that PCA may provide

misleading information for cross-sectional structure on high-dimensional data with dependent sample observations. This raises a natural question: *Can we trust PCA on high-dimensional data with general dependent observations?* Zhang, Pan, and Gao [39] contribute to this topic but with a focus on the difference between stationary and non-stationary unit root time series.

This paper is to work on a general sample structure as well as cross-sectional structure. The model under study is called a *separable covariance model*, which takes the form $\mathbf{\Omega}^{1/2}\mathbf{Z}\mathbf{\Sigma}^{1/2}$ with $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ being positive semi-definite matrices and $\mathbf{Z}$ possessing independent entries satisfying some moment assumptions. In this model, $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ describe the cross-sectional dependence and the sample directional dependence, respectively. Zhang [42] investigates the empirical spectral distribution (ESD) of the sample covariance matrix for the separable covariance model. This paper will study its largest spiked empirical eigenvalues for the separable covariance model. The results show that the spiked empirical eigenvalues are influenced by both the cross-sectional dependence and the dependent sample structure. In fact, when the effective rank of $\mathbf{\Omega}$ (see Fan *et al* [16] and Han and Liu [19]) defined by $r^*(\mathbf{\Omega}) = \text{Tr}(\mathbf{\Omega}) / \|\mathbf{\Omega}\|_2$ tends to infinity, our theoretical results quantify the influence of dependent sample observations on spiked empirical eigenvalues. These results suggest that implementing PCA or factor analysis by using popular methods for high-dimensional dependent data settings may lead to an incorrect conclusion for certain non-stationary models. As an additional contribution, we propose a test statistic to distinguish the unit root (with temporal dependence) from (nonstationary) factor structure (with cross-sectional dependence), based on the established properties of spiked empirical eigenvalues. In an empirical analysis, we investigate two popular data sets from economics and actuarial science, respectively. By applying the proposed test statistic to some determinants of healthcare expenditure in OECD countries, we show that a panel data model with interactive effects for this dataset is suitable. Moreover, the proposed test also justifies the feasibility of PCA for mortality data that indicate non-stationary temporal dependency.

The paper is organized as follows. Motivations from real data and simulated data are presented in Section 2. Section 3 introduces the data structure under study and some preliminary analysis. Asymptotic properties for spiked empirical eigenvalues are established in Section 4. An asymptotic theory for a test statistic, which is applied to high-dimensional time series, is proposed in Section 5. Section 6 illustrates the finite sample performance of the proposed test statistic based on some simulated data. Two empirical applications on OECD data and mortality data are discussed in Section 7. Conclusions and a discussion are provided in Section 8. All proofs are given in the Appendix.

2 Motivation

This section demonstrates human mortality data that possess cross-sectional dependence and time serial dependence simultaneously. Meanwhile, through a set of simulated data, we illustrate incorrect inference on cross-sectional dependence based on classical PCA.

2.1 Age-Specific Mortality Data

Annual age-specific death rates for the entire United States (U.S.) population are available for the years 1933 to 2016 in the Human Mortality Database (HMD). This dataset takes the form of matrix data, which consists of death rates for various ages from infants up to age 90 annually during a time period. For this matrix data, the dimension is the number of ages ($n = 91$) under study, while the sample size is the length of time observations ($T = 84$).

Figure 1 plots the time observations of the log central death rates for each age, while Figure 2 provides the age-observations of the log central death for each year. It follows from Figure 1 that time series for most ages are non-stationary, as they tend to decrease with time. Similar smooth curves in Figure 2 reflect that strong relationships (cross-sectional dependence) exist in the death rates at different ages.

Forecasting accuracy on mortality data is important for the management of pension funds and pricing of life insurance in actuarial science. Many existing mortality forecasting models follow the framework of classical factor models, such as the Lee-Carter model and its variants. As mortality data are high-dimensional time series, the Lee-Carter model, proposed in [25], utilizes PCA to extract one common feature (i.e. the first principal component) on a representation of the original data, and then establishes a forecasting model on the extracted feature. Although the Lee-Carter model performs relatively perfect in mortality forecasting, an interesting question that most statisticians commonly raise is how to justify PCA on mortality data, as mortality data are highly dependent, especially with a non-stationary tendency. In other words, *can we recover common features accurately from PCA with dependent sample observations?* Based on our theoretical results developed in Section 5, we will answer this question in Section 7.

Figure 1: The log central death rates for ages 0 to 90+ from year 1933 to year 2016

Figure 2: The log central death rates for years from 1933 to 2016 from age 0 to age 90+

2.2 Misleading PCA on Simulated Data

As is well known, factor model is equivalent to PCA to some extent (See [15]). PCA is then used to estimate common factors and factor loadings popular in the literature. In this section, we show that PCA provides misleading inference on the number of true factors in a simulated factor model with dependent sample observations.

Consider the factor model as follows:

$$y_{it} = \ell_{i1}f_{1t} + \ell_{i2}f_{2t} + \varepsilon_{it} = \boldsymbol{\ell}_i^* \mathbf{f}_t + \varepsilon_{it}, \quad i = 1, 2, \dots, n; \quad t = 1, 2, \dots, T, \quad (2.1)$$

where $\mathbf{f}_t = (f_{1t}, f_{2t})^*$ are two common factors, $\boldsymbol{\ell}_i = (\ell_{i1}, \ell_{i2})^*$ are the corresponding factor loadings, and ε_{it} is the error component, in which the symbol “ $*$ ” denotes the conventional conjugate transpose.

Common factors and error components follow the additional time series models below:

$$\mathbf{f}_t = \rho \mathbf{f}_{t-1} + \boldsymbol{\eta}_t \quad \text{and} \quad \varepsilon_{it} = \gamma \varepsilon_{i,t-1} + \xi_{it}, \quad (2.2)$$

where $\{\boldsymbol{\eta}_t : t = 1, 2, \dots, T\}$ are independent and identically distributed (i.i.d.) two-dimensional standard normal random vectors, $\{\xi_{it} : i = 1, 2, \dots, N; t = 1, 2, \dots, T\}$ are also i.i.d $\mathcal{N}(0, 1)$, and ρ and γ are two constants to take different values in our simulation.

As we know, common methods for estimating the number of factors in a factor model utilize information on spiked empirical eigenvalues, which are the largest leading eigenvalues of the sample covariance matrices. We will show eigenvalues of the sample covariance matrix $\mathbf{S} = \frac{1}{n} \mathbf{Y} \mathbf{Y}^*$ with $\mathbf{Y} = (y_{it})_{i,t=1}^{n,T}$ under two scenarios, respectively.

Scenario 1: No true common factors.

In this case, the factor loadings are generated as $\boldsymbol{\ell}_i = (0, 0)^*$. When the original data follows an AR(1) model ($\gamma = 0.2$), Figures 3 and 4 provide all eigenvalues of the sample covariance matrix as $(T, n) = (20, 40)$ and $(T, n) = (40, 20)$, respectively. There are no spiked eigenvalues in view of these graphs, which correctly reflect the fact that there are no common factors in the original data. However, as the data observations are non-stationary ($\gamma = 1$), Figures 5 and 6 show that there is one spiked eigenvalue from the sample covariance matrix, while the true number of common factors is 0.

Scenario 2: Two true common factors.

In this scenario, the factor loadings $\boldsymbol{\ell}_i$ are generated from a two-dimensional standard normal distribution. When ρ and γ are less than 1, y_{it} are stationary. Figures 7 and 8 show that there are 2 spiked largest eigenvalues from the sample covariance matrix under two high-dimensional cases, i.e. $(T, n) = (20, 40)$ and $(T, n) = (40, 20)$, respectively. They can accurately reflect the true number of factors in the factor model. However, as the residual (ε_{it}) is non-stationary ($\gamma = 1$), Figures 9 and 10 demonstrate one large spiked eigenvalue, which is inconsistent with the truth of two common factors in the factor model. Moreover, Figures 11 and 12 reveal a similar phenomenon when the common factors are non-stationary ($\rho = 1$). If ρ and γ are both equal to 1, Figures 13 and 14 also illustrate only 1 spiked eigenvalue from PCA under high-dimensional settings.

Figure 3: $T = 20, n = 40, \gamma = 0.2$

Figure 4: $T = 40, n = 20, \gamma = 0.2$

Figure 5: $T = 20, n = 40, \gamma = 1$

Figure 6: $T = 40, n = 20, \gamma = 1$

This example demonstrates that PCA may not be accurately applicable to high-dimensional data with dependent sample observations. After establishing the theoretical results, some other simulation examples will be further studied in Section 6.

3 High-Dimensional Separable Covariance Model

Consider an n -dimensional random vector $\mathbf{y}$ with observations $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$. Pool all observations together into a $T \times n$ matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)^*$. The data matrix $\mathbf{Y}$ has the structure

$$\mathbf{Y} = \mathbf{\Gamma} \mathbf{X} \mathbf{\Omega}^{1/2}, \quad (3.1)$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n) = (x_{ij})_{(T+L) \times n}$ is a $(T+L) \times n$ random matrix with i.i.d. elements; $\mathbf{\Sigma} = \mathbf{\Gamma} \mathbf{\Gamma}^*$ and $\mathbf{\Omega}$ are $T \times T$ and $n \times n$ deterministic non-negative definite matrices, respectively. Here $\mathbf{\Gamma}$ is a $T \times (T+L)$ deterministic matrix. The matrix $\mathbf{\Gamma}$ describes dependence among sample observations, while the matrix $\mathbf{\Omega}$ measures the cross-sectional dependence for $\mathbf{y}$ under study. Under this setting,

Figure 13: $T = 20, n = 40, \rho = 1, \gamma = 1$

Figure 14: $T = 40, n = 20, \rho = 1, \gamma = 1$

the sample covariance matrix of $\mathbf{y}$ can be expressed as $\mathbf{\Gamma}\mathbf{X}\mathbf{\Omega}\mathbf{X}^*\mathbf{\Gamma}^*$. It is also called a separable covariance matrix.

We are interested in the largest spiked eigenvalues of matrix $\mathbf{\Omega}$, which describes the cross-sectional dependence. In the classical procedure of PCA, spiked empirical eigenvalues from sample covariance matrix $\mathbf{\Gamma}\mathbf{X}\mathbf{\Omega}\mathbf{X}^*\mathbf{\Gamma}^*$ are utilized to approximate those of the matrix $\mathbf{\Omega}$. In this paper, we investigate the spiked empirical eigenvalues from an innovative view: how do the spiked eigenvalues of the matrix $\mathbf{\Sigma}$ (due to the dependent sample) affect the spiked sample eigenvalues? To this end, we do not impose any spiked structures on the matrix $\mathbf{\Omega}$, which can be either spiked or non-spiked.

We assume spikiness of the matrix $\mathbf{\Sigma}$ through the following decomposition. Let the spectral decomposition of $\mathbf{\Gamma}$ be $\mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{U}$, where $\mathbf{V}$ and $\mathbf{U}$ are $T \times T$ and $T \times (T + L)$ orthogonal matrices, respectively, ($\mathbf{V}\mathbf{V}^* = \mathbf{U}\mathbf{U}^* = \mathbf{I}$), and $\mathbf{\Lambda}$ is a diagonal matrix composed by the descent-ordered eigenvalues of $\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Gamma}^*$. Moreover, we write $\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_S & 0 \\ 0 & \mathbf{\Lambda}_P \end{pmatrix}$, where $\mathbf{\Lambda}_S = \text{diag}(\mu_1, \dots, \mu_K)$, $\mathbf{\Lambda}_P = \text{diag}(\mu_{K+1}, \dots, \mu_T)$, and $\mu_1, \dots, \mu_K$ are referred to the spiked eigenvalues that are significantly larger than the rest. In addition, we write $\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{pmatrix}$ and $\mathbf{\Sigma}_2 = \mathbf{U}_2^*\mathbf{\Lambda}_P\mathbf{U}_2$.

Our aim is to study the asymptotic behaviors of the largest K spiked eigenvalues of the sample covariance matrix $\mathbf{\Gamma}\mathbf{X}\mathbf{\Omega}\mathbf{X}^*\mathbf{\Gamma}^*$ under the setting of $\mathbf{\Sigma}$ being spiked. The formulation of $\mathbf{\Gamma}\mathbf{X}\mathbf{\Omega}\mathbf{X}^*\mathbf{\Gamma}^*$ is generally enough to cover some classical models commonly studied in the literature. When $\mathbf{\Sigma} = \mathbf{I}$, the original data $\mathbf{y}$ can possess cross-sectional dependence but independent sample observations. A classical example of measuring cross-sectional dependence is the factor model

$$\mathbf{Y} = \mathbf{F}\mathbf{\Lambda}_1 + \mathbf{Z}\mathbf{W} = (\mathbf{F}, \mathbf{Z}) \begin{pmatrix} \mathbf{\Lambda}_1 \\ \mathbf{W} \end{pmatrix}, \quad (3.2)$$

where $\mathbf{\Lambda}_1$ is the $r \times s$ factor loading, $\mathbf{F}$ is the corresponding $T \times r$ factor, $\mathbf{W}$ is $(n - r) \times s$ matrix, and $\mathbf{Z}$ is the idiosyncratic noise matrix. A common assumption is that the singular values of the factor

component $\mathbf{F}\mathbf{\Lambda}_1$ are significantly larger than those of the noise component. $\mathbf{Y}\mathbf{Y}^*$ can be written as a separable covariance model by setting the $n \times n$ matrix $\mathbf{\Omega} = \begin{pmatrix} \mathbf{\Lambda}_1\mathbf{\Lambda}_1^* & \mathbf{\Lambda}_1\mathbf{W}^* \\ \mathbf{W}\mathbf{\Lambda}_1^* & \mathbf{W}\mathbf{W}^* \end{pmatrix}$.

Meanwhile, when $\mathbf{\Omega} = \mathbf{I}$, the original data can possess dependent sample observations but are without cross-sections. A commonly used example is a linear process

$$y_{it} = \sum_{s=0}^L b_s x_{i,t-s}, \quad (3.3)$$

where $\{b_s : s = 0, 1, \dots, L\}$ are the coefficients satisfying some regular conditions and $x_{i,t}$ are i.i.d. random variables. It is easy to check that the model (3.3) can be written into a separable covariance model with $\mathbf{\Gamma} = \mathbf{W}$, where $\mathbf{W}$ is a $T \times (T + L)$ matrix with entries $W_{ij} = b_{L-j+i}I\{0 \leq j - i \leq L\}$.

It is worth mentioning that we will provide a non-stationary linear process with cross-sectional dependence in Section 4, which also belongs to our class of separable covariance models.

4 Asymptotic Properties for Largest Eigenvalues

This section will establish some asymptotic properties for the largest spiked empirical eigenvalues. First, we make the following assumptions.

Assumption 1 (Moment Conditions). $\{x_{ij} : i = 1, \dots, T + L, j = 1, \dots, n\}$ are i.i.d random variables such that $\mathbb{E}x_{ij} = 0$. $\mathbb{E}|\sqrt{n}x_{ij}|^2 = 1$ and $\mathbb{E}|\sqrt{n}x_{ij}|^4 = \gamma_4 < \infty$.

Assumption 2 (Dependent Sample Structure). $\alpha_{\mathcal{L}} = \mu_K = \dots = \mu_{K-n_{\mathcal{L}}} < \alpha_{\mathcal{L}-1} = \mu_{K-n_{\mathcal{L}}+1} \dots < \alpha_1 = \mu_{n_1} = \dots = \mu_1$, where $n_1, \dots, n_{\mathcal{L}}$ are finite. Moreover, there exists a small constant $c > 0$ such that $\alpha_{i-1} - \alpha_i \geq c\alpha_i$ for $i = 1, 2, \dots, \mathcal{L}$ and $\mu_K - \mu_{K+1} \geq c\mu_K$.

Assumption 3 (Cross-Sectional Structure). The matrix $\mathbf{\Omega}$ is non-negative definite and its effective rank $r^*(\mathbf{\Omega}) = \frac{\text{tr}(\mathbf{\Omega})}{\|\mathbf{\Omega}\|_2} \rightarrow \infty$, where $\|\mathbf{\Omega}\|_2$ means the spectral norm.

Assumption 4 (Spiked Dependent Sample Structure). The spiked eigenvalues of the population covariance matrix are much larger than the rest of the eigenvalues. Precisely speaking, for $\forall \varepsilon > 0$, there is K_ε , independent of n and T , such that when n and T are large enough,

$$\frac{\sum_{i=K_\varepsilon}^T \mu_i}{\mu_K} < \frac{\varepsilon}{2}. \quad (4.1)$$

Remark 1. Note that Assumptions 2 and 4 impose a spiked structure on $\mathbf{\Sigma}$ while Assumption 3 could impose either a spiked or a non-spiked structure on $\mathbf{\Omega}$. This is consistent with the aim of this paper to investigate the effect caused by dependent sample observations on spiked sample eigenvalues.

When $\mu_i = i^{-1-\sigma}$ and $\sigma > 0$ one can find that Assumption 4 holds. Moreover, Section 4.2 below shows that Assumption 4 holds in the unit root setting. In addition, define a near unit root model of the form:

$$y_{it} = \rho y_{i,t-1} + \sum_{h=0}^L b_h z_{i,t-h}, \quad (4.2)$$

where $T(1 - \rho)$ is bounded as T goes to infinity. It can also be verified that Assumption 4 holds in such models. Also, heterogeneous high-dimensional time series models can also be covered if the corresponding variances satisfy Assumption 4.

Denote the i -th largest sample eigenvalue of $\mathbf{\Gamma X \Omega X^* \Gamma^*}$ by λ_i . The following theorem establishes the asymptotic joint distribution of the largest spiked eigenvalues.

Theorem 1. Suppose that Assumptions 1-4 hold. Set $m_i = \sum_{j=1}^{i-1} n_j$, for all $i = 1, 2, \dots, \mathcal{L}$. Then

$$\frac{n}{\mu_i \sqrt{\text{tr}(\mathbf{\Omega}^2)}} \left(\lambda_{m_i+1} - \mu_i \frac{\text{tr} \mathbf{\Omega}}{n}, \lambda_{m_i+2} - \mu_i \frac{\text{tr} \mathbf{\Omega}}{n}, \dots, \lambda_{m_i+n_i} - \mu_i \frac{\text{tr} \mathbf{\Omega}}{n} \right) \xrightarrow{d} \mathcal{R}_i, \quad (4.3)$$

where $\mathcal{R}_i$ are the eigenvalues of an $n_i \times n_i$ matrix $\mathbf{R}_i$ with the Gaussian elements, $\mathbb{E} \mathbf{R}_i = 0$, the covariance of the $(\mathbf{R}_i)_{k_1, l_1}$, and $(\mathbf{R}_i)_{k_2, l_2}$ is

$$\lim_{n \rightarrow \infty} \frac{n^2}{\text{tr}(\mathbf{\Omega}^2)} \times \text{Cov}(\mathbf{u}_{m_i+k_1}^* \mathbf{X \Omega X^*} \mathbf{u}_{m_i+l_1}, \mathbf{u}_{m_i+k_2}^* \mathbf{X \Omega X^*} \mathbf{u}_{m_i+l_2}). \quad (4.4)$$

Here the limit of (4.4) is bounded.

Moreover, if $\frac{\mu_i}{\mu_j} \geq c > 1$, λ_{m_i+f} and λ_{m_j+g} are asymptotically independent, where $1 \leq f \leq n_i$ and $1 \leq g \leq n_j$. In particular, when $n_i = 1$ for all $i = 1, \dots, K$ we have

$$\left(\frac{\lambda_1 - \mu_1 \frac{\text{tr} \mathbf{\Omega}}{n}}{\mu_1 \sqrt{\text{var}(\mathbf{u}_1^* \mathbf{X \Omega X^*} \mathbf{u}_1)}}, \frac{\lambda_2 - \mu_2 \frac{\text{tr} \mathbf{\Omega}}{n}}{\mu_2 \sqrt{\text{var}(\mathbf{u}_2^* \mathbf{X \Omega X^*} \mathbf{u}_2)}}, \dots, \frac{\lambda_K - \mu_K \frac{\text{tr} \mathbf{\Omega}}{n}}{\mu_K \sqrt{\text{var}(\mathbf{u}_K^* \mathbf{X \Omega X^*} \mathbf{u}_K)}} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_K).$$

Remark 2. Note that

$$\sqrt{\frac{\text{tr}(\mathbf{\Omega})}{\|\mathbf{\Omega}\|_2}} = \frac{\text{tr}(\mathbf{\Omega})}{\sqrt{\|\mathbf{\Omega}\|_2 \text{tr}(\mathbf{\Omega})}} \leq \frac{\text{tr}(\mathbf{\Omega})}{\sqrt{\text{tr}(\mathbf{\Omega}^2)}} \leq \frac{\text{tr}(\mathbf{\Omega})}{\|\mathbf{\Omega}\|_2}. \quad (4.5)$$

From this and Assumption 3, we find that the standard deviation of λ_i has a smaller order than the mean of λ_i . So the sample eigenvalues $\{\lambda_i, i \leq K\}$ have the same order as $\mu_i \frac{\text{tr} \mathbf{\Omega}}{n}$ in probability. This indicates that the sample eigenvalues $\lambda_1, \dots, \lambda_K$ are spiked under this case no matter whether $\mathbf{\Omega}$ has spiked eigenvalues or not. This phenomenon suggests that PCA may reflect inaccurate information on the cross-sectional structure $\mathbf{\Omega}$ due to the dependent sample observations. This is in contrast to the results [7] for the independent sample observations, which establish one-to-one correspondence between the sample spiked eigenvalues and the population spiked eigenvalues due to the cross-sectional structure $\mathbf{\Omega}$.

The following two propositions further investigate the relations between the leading sample eigenvalues and the eigenvalues of $\mathbf{\Sigma}$ due to the dependent sample observations.

Proposition 1. Under the conditions of Theorem 1, there exists a positive constant c such that $\liminf_{T \rightarrow \infty} \frac{\mu_1}{\text{tr}(\mathbf{\Gamma \Gamma^*})} > c$ and

$$\lim_{n, T \rightarrow \infty} P \left(\frac{\lambda_1}{\text{tr}(\mathbf{\Gamma X \Omega X^* \Gamma^*})} > c \right) = 1. \quad (4.6)$$

Moreover, when $1 \leq i < K$,

$$\frac{\lambda_i}{\lambda_{i+1}} - \frac{\mu_i}{\mu_{i+1}} \rightarrow 0 \text{ in probability as } n, T \rightarrow \infty. \quad (4.7)$$

Proposition 2. *Let the conditions of Theorem 1 hold. For $1 \leq i < K-1$, if $\min \left\{ \frac{\mu_i}{\mu_{i+1}}, \frac{\mu_{i+1}}{\mu_{i+2}} \right\} \geq c > 1$, then $\frac{\frac{\lambda_i - \lambda_{i+1}}{\mu_i - \mu_{i+1}}}{\frac{\lambda_{i+2} - \lambda_{i+1}}{\mu_{i+2} - \mu_{i+1}}} = \frac{\lambda_i \frac{\mu_{i+1}}{\mu_i} - \lambda_{i+1}}{\lambda_{i+2} \frac{\mu_{i+1}}{\mu_{i+2}} - \lambda_{i+1}}$ has the same asymptotic distribution as $\frac{v_1 - v_2}{v_3 - v_2}$, where $\{v_i : 1 \leq i \leq 3\}$ are i.i.d standard normal random variables.*

Remark 3. *Proposition 1 shows that the ratio of the neighboring spiked empirical eigenvalues approximates that of the spiked eigenvalues from the dependent sample structure. A central limit theorem is provided for the ratio statistic constructed from the spiked empirical eigenvalues in Proposition 2.*

5 Inference on High-Dimensional Time Series

This section explores the applications of the established asymptotic results to high-dimensional time series. To this end, we first show that applying existing statistical methods of determining the number of factors in factor models to our data structures may result in an incorrect conclusion. In other words, implementing PCA/factor analysis on our data structure will lead to a misleading conclusion. This means that we must rule out such models before implementing PCA. However, unit root models turn out to satisfy our model assumptions in Section 2, as illustrated below. Hence, it is desirable to distinguish a high-dimensional unit root structure from a high-dimensional (non-stationary) factor model. We propose a test statistic to serve this purpose.

5.1 Implementing Factor Analysis on Our Model

Proposition 1 implies that the largest sample eigenvalue has the same order as the sum of all eigenvalues. The largest eigenvalue is so large that the methods in [2] and [3] would both estimate the number of factors to be those larger than zero even though there is no factor in our model. Similarly, the relation between λ_i and λ_{i+1} leads to the method of [29] estimating the number of factors to be that larger than zero even though there is no factor in our model. We examine this one by one below.

The method in [29] is based on the difference between the i -th largest eigenvalue and the $i+1$ -th one. In a few words, the idea of [29] is that if there is no factor, for any $i \geq 1$, the difference between the i -th largest eigenvalue and the $i+1$ -th one should be very small. Recalling (4.7), we find that the difference between the i -th largest eigenvalue and the $i+1$ -th one in our model can be large when $\frac{\mu_i}{\mu_{i+1}} > 1$. In other words, the method in [29] would obtain a non-zero estimate for the number of factors in our model when $\frac{\mu_i}{\mu_{i+1}} > 1$.

The method in [2] is based on the ratio between two successive largest eigenvalues. It defines a mock eigenvalue $\lambda_0 = \frac{\sum_{i=1}^{\min\{n, T\}} \lambda_i}{\ln(\min\{n, T\})}$. Then the estimator is

$$\tilde{k}_{ER} = \max_{0 \leq k \leq k_{max}} \frac{\lambda_k}{\lambda_{k+1}}. \quad (5.1)$$

Note that λ_0 has a smaller order than the trace of the sample covariance matrix. Then, the method of [2] implies that if there is no factor, the largest eigenvalue should have a smaller order than the trace of the sample covariance matrix. (4.6) shows that the largest eigenvalue in our model has the same

order as the trace of the sample covariance matrix. In other words, the method in [2] would obtain a non-zero estimate for the number of factors in our model. This is also reported in Proposition 3 below.

The methods in [3] is based on penalty functions. Briefly, the idea of [3] is that if there is no factor, the largest eigenvalue should be smaller than the penalty function. The criterion has the form:

$$PC(k) = \min_{\Lambda} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (X_{it} - \lambda_i^{k'} \hat{F}_t^k)^2 + kg(n, T), \quad (5.2)$$

where $g(n, T)$ satisfies some properties in [3]. However, we find that all the penalty functions have a smaller order than the trace of the sample covariance matrix. Recalling (4.6), the methods in [3] would obtain a non-zero estimate for the number of factors in our model.

5.2 Unit Root Models Satisfying Assumption 4

This subsection verifies that certain unit root models satisfy Assumption 4. Consider a unit root model of the form:

$$y_{it} = y_{i,t-1} + \sum_{h=0}^L b_h z_{i,t-h}, \quad (5.3)$$

where $1 \leq i \leq n$, $1 \leq t \leq T$ and L can be finite or infinite. Here

$$z_{it} = \sum_{s=1}^n \Upsilon_{is} x_{st}, \quad (5.4)$$

where $1 \leq i \leq n$ and $1 - L \leq t \leq T$.

Let $\mathbf{Y} = (y_{it})$ be an $n \times T$ matrix and $\bar{\mathbf{Y}}$ be an $n \times T$ matrix with all entries of the i -th row being $\frac{\sum_{t=1}^T y_{it}}{T}$. We next specify some conditions so that Theorem 1 can be applied to the sample covariance matrix: $(\mathbf{Y} - \bar{\mathbf{Y}})^*(\mathbf{Y} - \bar{\mathbf{Y}})$.

Assumption 5 (Moment Conditions). $\{x_{it}: i = 1, \dots, n, t = 1 - L, \dots, T\}$ are independent random variables such that $\mathbb{E}x_{it} = 0$. $\mathbb{E}|\sqrt{n}x_{it}|^2 = 1$ and $\mathbb{E}|\sqrt{n}x_{it}|^4 = \gamma_4 < \infty$. Write $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$.

Assumption 6 (Cross-sectional Structure). $\Omega = \Upsilon^* \Upsilon$ satisfies Assumption 3 with the $n \times n$ matrix $\Upsilon = (\Upsilon_{is})$.

Assumption 7 (Serial Correlation). The coefficients $\{b_i\}_{i=0}^L$ in (5.3) satisfy $\sum_{i=0}^L i|b_i| < \infty$ and $\sum_{i=0}^L b_i \neq 0$.

We next define $\mathbf{W}$ to be an $T \times (T + L)$ matrix with entries $W_{ij} = b_{L-j+i} I\{0 \leq j - i \leq L\}$. Let $\mathbf{C} = (C_{ij})_{1 \leq i, j \leq T}$ be a $T \times T$ lower triangular matrix with

$$C_{ij} = 0 \text{ for } j > i \text{ and } C_{ij} = 1 \text{ for } 1 \leq j \leq i. \quad (5.5)$$

Write $\mathbf{H} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^*}{T}$, where the $T \times 1$ vector $\mathbf{1}$ consists of all one. Let $\Gamma = \mathbf{H}\mathbf{C}\mathbf{W}$ and $\mu_1 \geq \mu_2 \geq \dots \geq \mu_T$ be the ordered eigenvalues of $\Gamma\Gamma^*$. With a simple calculation, the sample covariance matrix can be expressed as

$$(\mathbf{Y} - \bar{\mathbf{Y}})^*(\mathbf{Y} - \bar{\mathbf{Y}}) = \mathbf{H}\mathbf{Y}^*\mathbf{Y}\mathbf{H}^* = \Gamma\mathbf{X}\Omega\mathbf{X}^*\Gamma^*. \quad (5.6)$$

We are now in a position to apply Theorem 1 to obtain the following result for model (5.3).

Theorem 2. *Let Assumptions 5-7 hold. Denoting the k -th largest eigenvalue of (5.6) by λ_k , for any fixed k , we have*

$$\frac{n}{\sqrt{2\text{tr}(\Omega^2)}} \left(\frac{\lambda_1 - \mu_1 \frac{\text{tr}\Omega}{n}}{\mu_1}, \frac{\lambda_2 - \mu_2 \frac{\text{tr}\Omega}{n}}{\mu_2}, \dots, \frac{\lambda_k - \mu_k \frac{\text{tr}\Omega}{n}}{\mu_k} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_k). \quad (5.7)$$

Propositions 1-2 hold as well for model (5.3).

Remark 4. *Theorem 2 is more general than the result in [39]. Assumption 6 (Assumption 3) here is significantly different from A3 in [39]. A3 in [39] assumes the largest eigenvalues due to the cross-sectional structure to be bounded and that the trace of the cross-sectional structure has a certain order. In contrast to [39], Assumption 6 here only assumes that the largest eigenvalues due to the cross-sectional structure have a smaller order than the trace. In consequence, Theorem 1 works for many more models, including unit root models.*

Furthermore, Theorem 1 holds for a more general random matrix $\mathbf{H}_1 \mathbf{Y}^* \mathbf{H}_2 \mathbf{Y} \mathbf{H}_1^*$, where $\mathbf{H}_1$ and $\mathbf{H}_2$ can denote some operations such as de-meaning and de-trending for different types of fixed terms in different models, if $\mathbf{Y}^* \mathbf{H}_2 \mathbf{Y}$ satisfies the Assumption 3 and $\mathbf{H}_1 \mathbf{C} \mathbf{W}^*$ satisfies Assumptions 2 and 4.

5.3 A New Test for Unit Root Models against Factor Models

In view of the above two sections, it is desirable to distinguish between unit root and (non-stationary) factor models. Here, one should note that when the factor model is stationary it seems relatively easy to find the differences between the unit root and factor model under consideration. Hence, we focus on a non-stationary factor model of the form:

$$M1 : y_{it} = \ell_i^* \mathbf{f}_t + e_{it}, \quad (5.8)$$

where $\mathbf{f}_t$ is a r -dimensional (r is fixed) vector, e_{it} is a stationary term and $\{\mathbf{f}_t\}_{t=1, \dots, T}$ are independent of $\{e_{it}\}_{i=1, \dots, n, t=1, \dots, T}$.

We then recall the unit root model discussed in Theorem 2 as follows:

$$M2 : y_{it} = y_{i,t-1} + \sum_{h=0}^L b_h z_{i,t-h} \quad (5.9)$$

for $1 \leq i \leq n$ and $1 \leq t \leq T$, where L can be finite or infinite, $z_{it} = \sum_{s=1}^n \Upsilon_{is} x_{st}$ and Assumptions 5-7 hold.

Consider the hypotheses:

$$H_0 : y_{it} \text{ follows } M2 \text{ versus } H_1 : y_{it} \text{ follows } M1. \quad (5.10)$$

Remark 5. *Note that Model M1 is equivalent to the following form:*

$$M3 : y_{it} = y_{i,t-1} + \ell_i^* (\mathbf{f}_t - \mathbf{f}_{t-1}) + e_{it} - e_{i,t-1} \triangleq y_{i,t-1} + \xi_{it}, \quad (5.11)$$

which seems to be similar to model M2. However, M3 is different from M2 in two aspects. At first, if $\mathbf{f}_t - \mathbf{f}_{t-1} \neq 0$, $\ell_i^*(\mathbf{f}_t - \mathbf{f}_{t-1})$ could lead to a strong cross-sectional dependence (strong factor) such that Assumption 6 is violated. Furthermore, even if $\mathbf{f}_t - \mathbf{f}_{t-1} = 0$, then $\xi_{it} = e_{it} - e_{i,t-1}$ does not necessarily satisfy $\sum_{i=0}^L b_i \neq 0$ in Assumption 7. Here, one should note that the term $\left(\sum_{i=0}^L b_i\right)^2$ contributes to the limit of the first few largest eigenvalues of the corresponding sample covariance matrix although the explicit expression of μ_i in Theorem 2 is not given here (one may refer to its asymptotic expression in Definition 1 and Lemma 1 in the Appendix).

As a result, the eigenvalues of these two models behave differently. When the number of factors in M1 is r , there is a significant drop from the r -th largest eigenvalue of M1 to its $(r+1)$ -th largest eigenvalue. In contrast, the ratio of the i -th largest eigenvalue of M2 to its $(i+1)$ -th largest eigenvalue is asymptotically equal to $(i+1)^2/i^2$ (see the proof of Proposition 3 in the Appendix) so that there is no significant drop between them. These observations are also reflected in (5.14) and (5.17) below.

We propose a new statistic for (5.10). From Proposition 2 one may consider using $\frac{\lambda_i \frac{\mu_{i+1}}{\mu_i} - \lambda_{i+1}}{\lambda_{i+2} \frac{\mu_{i+1}}{\mu_{i+2}} - \lambda_{i+1}}$ as a statistic since it has an asymptotic distribution under H_0 . However, $\{\mu_i\}$, the eigenvalues of Ω are unknown and difficult to estimate. Fortunately, the ratio of $\frac{\mu_{i+1}}{\mu_i}$ can be consistently estimated by $\frac{\bar{\mu}_{i+1}}{\bar{\mu}_i}$ where

$$\bar{\mu}_i = \frac{1}{2 \left(1 + \cos \left(\frac{(T-i)\pi}{T}\right)\right)}. \quad (5.12)$$

Hence, a new statistic for (5.10) can be proposed as follows

$$T_{uf} = \frac{\lambda_1 \frac{\bar{\mu}_2}{\bar{\mu}_1} - \lambda_2}{\lambda_2 - \lambda_3 \frac{\bar{\mu}_2}{\bar{\mu}_3}}. \quad (5.13)$$

Proposition 3. Under the conditions of Theorem 2, for any fixed k ,

$$\lim_{n,T \rightarrow \infty} P \left(\tilde{k}_{ER} = \max_{0 \leq i \leq k} \left\{ \frac{\lambda_i}{\lambda_{i+1}} \right\} = 1 \right) = 1. \quad (5.14)$$

Furthermore, when

$$\lim_{n,T \rightarrow \infty} \frac{\text{tr}(\Omega)}{T \sqrt{\text{tr}(\Omega^2)}} = 0, \quad (5.15)$$

the statistic T_{uf} has the same asymptotic distribution as $\frac{v_1 - v_2}{v_2 - v_3}$, where $\{v_i : 1 \leq i \leq 3\}$ are i.i.d standard normal variables.

Remark 6. Note that $\frac{\text{tr}(\Omega)}{\sqrt{\text{tr}(\Omega^2)}} \leq \sqrt{n}$. So when $\frac{\sqrt{n}}{T} \rightarrow 0$, (5.15) holds.

Equation (5.14) implies that using $\tilde{k}_{ER}$ in [2] may mistakenly consider a unit root model as a single factor model. However, using the statistic T_{uf} can distinguish between them.

Assumption 8 (Single Factor). Let y_{it} follow the model M1 and $r = 1$. Moreover, $\mathbf{E}$ is the matrix with elements e_{it} and $\bar{f}_T = \frac{1}{T} \sum_{i=1}^T f_t$ such that

$$\frac{\sum_{i=1}^n \sum_{t=1}^T l_i^2 (f_t - \bar{f}_T)^2}{\|\mathbf{E}\|_2^2} \rightarrow \infty \text{ in probability as } n, T \rightarrow \infty. \quad (5.16)$$

Assumption 8 leads to the following relation:

$$\frac{\lambda_1}{\lambda_2} \rightarrow \infty \text{ in probability as } n, T \rightarrow \infty. \quad (5.17)$$

This ensures the power of the statistic T_{uf} specified below.

Proposition 4. *Suppose y_{it} satisfies Assumption 8. Then we have*

$$T_{uf} \rightarrow \infty \text{ in probability as } n, T \rightarrow \infty. \quad (5.18)$$

Remark 7. *It's easy to verify Assumption 8 in single factor models. In fact, either Assumptions 1-2 in [29] or Assumptions A-C in [3] ensure Assumption 8, so do Assumptions A-C in [2]. Moreover, Assumption 8 contains many other cases, such as $f_t = t^2$ or $f_t = \sin(\pi t)$. This means our test can be applicable to not only linear factor models with unit root, but also the general nonstationary factor models that are driven by nonlinear deterministic trends.*

5.4 Further Thoughts on Panel Data Structures

Along with [39], we have further thoughts on distinguishing four kinds of panel data structures:

- (1) stationary and weak cross-sectional dependence;
- (2) stationary and strong cross-sectional dependence;
- (3) unit root and weak cross-sectional dependence;
- (4) unit root and strong cross-sectional dependence.

Here, by strong cross-sectional dependence we mean that its effective rank $r^*(\mathbf{\Omega}) = \frac{\text{tr}(\mathbf{\Omega})}{\|\mathbf{\Omega}\|_2} \rightarrow c > 0$ while weak cross-sectional dependence implies that its effective rank $r^*(\mathbf{\Omega}) \rightarrow +\infty$. Then, this paper and [39] pave the way to classify the panel data. Theorem 2.3 of [39] shows that when the data belongs to the first kind, the largest eigenvalue of sample covariance matrix has a smaller order than the trace. On the other hand, the largest eigenvalues from the three other types of data have the same order as the trace. Thus, we can distinguish the first type, stationary and weak cross-sectional dependence, from the others.

For the remaining three cases, we consider using PCA. Since PCA is a linear combination of data on the cross-section we believe it has the same time-dependence as the initial data. In other words, from the first PC we may tell the difference between the second type stationary structure and the remaining two non-stationary structures since there are plenty of methods available for the univariate variable. Thus, we can distinguish the second case from two others. Finally, we can use T_{uf} to distinguish the third case from the fourth case.

6 Simulations

This section conducts some simulations for Propositions 3 and 4, which also demonstrates the finite sample performance of the proposed test statistic.

6.1 The Simulation for Proposition 3

At first we compute the critical value by simulating $\frac{v_1-v_2}{v_2-v_3}$ where $\{v_i : 1 \leq i \leq 3\}$ are i.i.d standard Gaussian random variables based on 500000 replications. The quantiles of $\frac{v_1-v_2}{v_2-v_3}$ are reported in Table 1. Then we can use a two-tailed test with the critical values -11.6549 and 10.4598. We can also use

Table 1: The quantiles of $\frac{v_1-v_2}{v_2-v_3}$ based on 500000 replications

2.5 percent quantile	5 percent quantile	95 percent quantile	97.5 percent quantile
-11.6549	-6.0392	4.9932	10.4598

a one-side test with the critical values $C_{5L} = -6.0392$ or $C_{95R} = 4.9932$.

We consider the following setting:

$$y_{it} = y_{i,t-1} + \psi z_{i,t-1} + z_{it},$$

where $\psi = 0.5$ and $\mathbf{\Omega} = (\Omega_{i,j}) = (0.3^{|i-j|})$. This data-generating process satisfies the unit root model discussed in Section 5.2.

The estimated sizes for the test statistic T_{uf} based on 1000 replications, different critical values, and different values of n and T are reported in Tables 2-3. Tables 2-3 show that T_{uf} has stable sizes with different critical values. We can choose -11.6549 and 10.4598 as the critical values of a two-tailed test.

We also calculate the proportion of $\tilde{k}_{ER} = 1$ with different values of the prescribed upper bound k in (5.14). Tables 4-5 show that (5.14) also works well, since the calculated proportion is approaching 1 as both n and T increase.

6.2 The Simulation for Proposition 4

Now we consider the single factor model:

$$y_{it} = l_i f_t + \sqrt{\theta} e_{it}. \quad (6.1)$$

We use the same error term e_{it} as in [2]: $e_{it} = \sqrt{\frac{1-\rho^2}{1+2J\beta}} \tilde{e}_{it}$

$$\tilde{e}_{it} = \rho \tilde{e}_{i,t-1} + \epsilon_{it} + \sum_{h=\max(i-J,1)}^{i-1} \beta \epsilon_{ht} + \sum_{i+1}^{h=\min(i+J,n)} \beta \epsilon_{ht}, \quad (6.2)$$

where ϵ_{it} and l_i are all drawn independently from $N(0, 1)$. We also use the most complicated setting of [2] which has both serially and cross-sectionally correlated errors: $\rho = 0.5, \beta = 0.2$ and $J = \max(10, n/20)$.

Since $\rho < 1$, we find that the error term is stationary. If f_t is also stationary, y_{it} is stationary. Then it will be very different from the unit root model, and there are too many methods to test stationary and unit root models. Hence, we focus on the case where f_t is non-stationary. We set

$f_t = f_{t-1} + \tilde{f}_t$, where $\tilde{f}_t$ is drawn independently from $N(0, 1)$. Then the power of T_{uf} based on 1000 replications, different θ , the critical values of the two-sided test and different values of n and T , are reported in Tables 6-7. The power results given in both Tables demonstrate that the proposed test works well numerically.

Table 2: The size results on T_{uf} based on 1000 replications

n	T	two-sided	C_{5L}	C_{95R}
20	20	0.069	0.064	0.052
20	40	0.065	0.068	0.051
20	60	0.067	0.059	0.055
20	80	0.074	0.062	0.063
20	100	0.064	0.060	0.060
20	200	0.055	0.064	0.054
40	20	0.062	0.069	0.063
40	40	0.070	0.061	0.062
40	60	0.055	0.063	0.057
40	80	0.052	0.068	0.047
40	100	0.052	0.054	0.056
40	200	0.060	0.059	0.046
60	20	0.051	0.045	0.051
60	40	0.048	0.051	0.048
60	60	0.047	0.050	0.047
60	80	0.055	0.057	0.056
60	100	0.050	0.043	0.055
60	200	0.053	0.050	0.057

7 Empirical Applications

This section conducts analysis on two sets of real data: one is an OECD healthcare expenditure dataset, while the other is U.S. mortality data, both of which are popular in health economics and actuarial science (as well as demography), respectively. We will apply the proposed test statistic to check whether the data possess a unit root structure or a factor model.

Table 3: The size results on T_{uf} based on 1000 replications

n	T	two-sided	C_{5L}	C_{95R}
80	20	0.062	0.052	0.064
80	40	0.059	0.047	0.051
80	60	0.051	0.051	0.057
80	80	0.054	0.059	0.048
80	100	0.047	0.058	0.047
80	200	0.054	0.044	0.055
100	20	0.058	0.048	0.050
100	40	0.057	0.061	0.042
100	60	0.035	0.046	0.044
100	80	0.051	0.058	0.044
100	100	0.051	0.040	0.054
100	200	0.053	0.059	0.039
200	20	0.048	0.040	0.059
200	40	0.058	0.055	0.045
200	60	0.046	0.043	0.048
200	80	0.046	0.050	0.044
200	100	0.055	0.048	0.057
200	200	0.066	0.046	0.063

Table 4: The proportion of $\tilde{k}_{ER} = 1$ based on 1000 replications

n	T	$k = 5$	$k = 10$	$k = 15$
20	20	0.720	0.712	0.701
20	40	0.733	0.732	0.732
20	60	0.749	0.749	0.749
20	80	0.770	0.770	0.770
20	100	0.754	0.754	0.754
20	200	0.766	0.766	0.766
40	20	0.816	0.815	0.815
40	40	0.841	0.841	0.841
40	60	0.835	0.835	0.835
40	80	0.835	0.835	0.835
40	100	0.824	0.824	0.824
40	200	0.838	0.838	0.838
60	20	0.874	0.874	0.874
60	40	0.870	0.870	0.870
60	60	0.902	0.902	0.902
60	80	0.897	0.897	0.897
60	100	0.877	0.877	0.877
60	200	0.897	0.897	0.897

Table 5: The proportion of $\tilde{k}_{ER} = 1$ based on 1000 replications

n	T	$k = 5$	$k = 10$	$k = 15$
80	20	0.920	0.920	0.920
80	40	0.933	0.933	0.933
80	60	0.909	0.909	0.909
80	80	0.914	0.914	0.914
80	100	0.925	0.925	0.925
80	200	0.914	0.914	0.914
100	20	0.936	0.936	0.936
100	40	0.950	0.950	0.950
100	60	0.931	0.931	0.931
100	80	0.937	0.937	0.937
100	100	0.946	0.946	0.946
100	200	0.945	0.945	0.945
200	20	0.981	0.981	0.981
200	40	0.986	0.986	0.986
200	60	0.989	0.989	0.989
200	80	0.986	0.986	0.986
200	100	0.988	0.988	0.988
200	200	0.982	0.982	0.982

Table 6: The power results on T_{uf} based on 1000 replications and a two-sided test

n	T	$\theta = 3$	$\theta = 1$	$\theta = 1/3$
20	20	0.087	0.232	0.582
20	40	0.114	0.380	0.725
20	60	0.147	0.485	0.843
20	80	0.197	0.605	0.896
20	100	0.249	0.634	0.953
20	200	0.455	0.868	0.995
40	20	0.116	0.295	0.679
40	40	0.173	0.474	0.848
40	60	0.260	0.637	0.946
40	80	0.293	0.715	0.974
40	100	0.389	0.768	0.980
40	200	0.620	0.932	0.999
60	20	0.082	0.289	0.710
60	40	0.122	0.454	0.848
60	60	0.151	0.554	0.925
60	80	0.223	0.653	0.958
60	100	0.282	0.764	0.979
60	200	0.509	0.934	0.999

Table 7: The power results on T_{uf} based on 1000 replications and a two-sided test

n	T	$\theta = 3$	$\theta = 1$	$\theta = 1/3$
80	20	0.077	0.267	0.678
80	40	0.129	0.448	0.854
80	60	0.170	0.544	0.939
80	80	0.206	0.687	0.961
80	100	0.285	0.762	0.982
80	200	0.537	0.925	1
100	20	0.061	0.288	0.689
100	40	0.095	0.437	0.866
100	60	0.157	0.603	0.968
100	80	0.208	0.709	0.980
100	100	0.287	0.806	0.985
100	200	0.599	0.969	1
200	20	0.041	0.238	0.674
200	40	0.085	0.492	0.894
200	60	0.192	0.680	0.978
200	80	0.327	0.794	0.995
200	100	0.442	0.883	0.999
200	200	0.761	0.991	1

7.1 OECD Health Expenditure Analysis

Healthcare expenditure in the OECD* varies substantially over time and across countries. As indicated in [8], from 1970 to 2004, per capita health expenditure has increased markedly in the OECD, with an annual average rate of 11.5%. A number of econometric models have been proposed in the literature to regress healthcare expenditure on certain determinants, as well as to capture the homogeneity and heterogeneity in healthcare expenditure among OECD countries by factor models. For example, [18] provides a panel data model with heterogeneous time trending and fixed effects.

In this empirical example, we apply our test statistic to the healthcare expenditure dataset and some of its determinants to check whether there is a factor structure in each of these random variables. This study also provides statistical support for some econometric models associated with factor structures on these popular random variables. We collect the annual data on 16 OECD countries ($n = 16$) from 1976 to 2010 ($T = 35$) for five variables from the OECD Health Statistics 2015 database, including per capita health care expenditure (HE), per capita gross domestic product (GDP), the proportions of the overall population aged below 15 and above 65 (DR_y and DR_o , respectively), and the proportion of government funding invested in the healthcare industry (GHE). Data plots for these variables are presented in Figures 15-19. It can be seen from Figure 15 that HE generally continues to increase from 1976 to 2010 for most countries. The temporal dynamics clearly demonstrate non-stationarity. Moreover, the phenomenon of similar temporal trends for all countries also illustrates the strong relations among these countries (i.e. cross-sectional dependence). In the same analysis on Figures 16-18, variables GDP , DR_y , and DR_o also possess strong cross-sectional dependence as well as non-stationary trend. However, Figure 19 shows that the variable GHE does not have obvious cross-sectional dependence and non-stationary trend. In the econometrics literature, the relation between HE and its determinants is described by the following model:

$$HE_{it} = \beta_{1i} \cdot GDP_{it} + \beta_{2i} \cdot DR_{y,it} + \beta_{3i} \cdot DR_{o,it} + \beta_4 \cdot GHE_{it} + u_{it}, \quad (7.1)$$

$$u_{it} = \boldsymbol{\lambda}_i^* \mathbf{F}_t + \varepsilon_{it}, \quad (7.2)$$

where $\beta_{ki}, k = 1, 2, 3, 4$ are unknown coefficients, $\boldsymbol{\lambda}_i$ is r -dimensional factor loading, $\mathbf{F}_t$ is the corresponding factor, and ε_{it} is the idiosyncratic error component.

Now we apply the proposed test statistic T_{uf} to each of these five variables (i.e. HE , GDP , DR_y , DR_o , GHE) as well as the estimated residual, $\hat{u}_{it}$, in the following procedure. Take HE as an example.

1. Calculate the eigenvalues $\lambda_1 \geq \lambda_2 \leq \dots \leq \lambda_{\min(n,p)}$ of the sample covariance matrix $\mathbf{B} = \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})^*$ for the random vector $\mathbf{x}_t = (HE_{1t}, HE_{2t}, \dots, HE_{nt})^*$, where $\bar{\mathbf{x}} = \left(\frac{1}{T} \sum_{t=1}^T HE_{1t}, \frac{1}{T} \sum_{t=1}^T HE_{2t}, \dots, \frac{1}{T} \sum_{t=1}^T HE_{nt} \right)^*$.
2. Choose a positive integer $k_{\max} = 10$ and calculate the estimator

$$\tilde{k}_{ER} := \max_{0 \leq i \leq k_{\max}} \frac{\lambda_i}{\lambda_{i+1}}. \quad (7.3)$$

*Organization for Economic Cooperation and Development

If $\tilde{k}_{ER} = 0$, the corresponding variable, x_{it} , does not have factor structure or unit root with probability 1; if $\tilde{k}_{ER} \geq 2$, x_{it} is not unit root with probability 1; if $\tilde{k} = 1$, we need to conduct the test in the following step.

3. Calculate the value of the statistic T_{uf} in Proposition 3 and check whether it falls into the 95% confidence interval $[-11.6549, 10.4598]$. If it falls into this interval, our conclusion is that x_{it} possesses unit root structure; otherwise it follows a factor model.

The estimated values of $\tilde{k}_{ER}$ and T_{uf} for these five random variables are listed in Table 8. Empirical results show that dependent variable HE , the regressor variables GDP and DR_y , and the residual u_{it} possess factor structures, while DR_o and GHE follow unit root structures. These results are consistent with the relevant literature using a factor structure to model healthcare expenditure (HE), such as [8] and [18].

Table 8: Test results for HE and its determinants						
	HE	GDP	DR_y	DR_o	GHE	Residual
$\tilde{k}_{ER}$	1	1	1	1	1	1
T_{uf}	-60.5999	76.7654	61.4004	8.9181	7.6524	-12.2605
Conclusion	False	False	False	True	True	False

Furthermore, we analyze the eigenvalues of the sample covariance matrices for HE , GDP , DR_y , DR_o , GHE , and the residual u_{it} , which are provided in Figures 21-26, respectively. As mentioned in Section 4, the empirical eigenvalues and population eigenvalues have large differences in statistical performance on high-dimensional data with dependent samples, which in turn results in inaccurate inference for the number of factors in the factor model with certain criteria based on sample covariance matrices, such as [2], [3], and [29]. Figures 21-24 show that there appears to be one factor in HE , GDP , DR_y , DR_o , GHE , and the residual u_{it} . The estimate $\tilde{k}_{ER}$ used in [2] is also provided in Table 8, which is also equal to 1 for HE , GDP , DR_y , DR_o , GHE , and the residual u_{it} . The test results show that this information is inaccurate for DR_o and GHE , since they follow a unit root structure rather than a factor model.

7.2 U.S. Mortality Data Analysis

Mortality forecasting is an important topic in various areas, such as demography, actuarial science, and government policymaking. Age-specific mortality data are high-dimensional time series, which consist of observations for the death rates for each age. Many existing mortality models follow the framework of factor models, such as the prominent Lee-Carter model proposed in [25] and its variants.

The mortality data of the U.S. used here are extracted from the HMD (14). For the years from 1933 to 2016, age-specific death rates are available annually for the entire U.S. population. Thus, we obtain a total of $T = 84$ yearly observations for each age. Moreover, the mortality data are available from age 0 to age 110+ for each year. As the mortality data for advanced ages are sparse and noisy

Figure 15: Observations of HE for 16 countries from 1976 to 2010

Figure 16: Observations of GDP for 16 countries from 1976 to 2010

Figure 17: Observations of DR_y for 16 countries from 1976 to 2010

Figure 18: Observations of DR_o for 16 countries from 1976 to 2010

Figure 19: Observations of GHE for 16 countries from 1976 to 2010

Figure 20: Observations of the residuals for 16 countries from 1976 to 2010

Figure 21: Eigenvalues of sample covariance matrix for HE

Figure 22: Eigenvalues of the sample covariance matrix for GDP

Figure 23: Eigenvalues of the sample covariance matrix for DR_y

Figure 24: Eigenvalues of the sample covariance matrix for DR_o

Figure 25: Eigenvalues of the sample covariance matrix for GHE Figure 26: Eigenvalues of the sample covariance matrix for the residuals

(mentioned in [25]), we ignore the mortality data for ages from 90 to 110+, and replace the death rates for age 90 with d_{90+}/E_{90+} , where d_{90+} is the number of deaths and E_{90+} is the number of exposures to risk for age 90 and above. Then the ages under study are ages from 0 to 90[†] and thus the dimension of this data is $n = 91$. As the time length $T = 84$ is comparable to the dimension $n = 91$, it belongs to high-dimensional time series. We use the logarithms of the death rates in our analysis, as the transformed data are not limited to positive values as the original death rates.

An interesting problem appears: whether the common factor extracted from the Lee-Carter model (i.e. factor model) describes cross-sectional dependence? To answer this question, we apply the proposed test statistic on this data. Based on the procedure outlined in Section 6.1, empirical results are provided in Table 9. We conclude that the U.S. mortality data follows a one-factor model, which supports the Lee-Carter model as a benchmark model in mortality forecasting. Moreover, Figure 27 illustrates empirical eigenvalues from the sample covariance matrix, which shows the largest eigenvalue is much larger than the rest. The estimate $\tilde{k}_{ER}$ in Table 9 indeed equals 1 for this data. Our test result also confirms that the estimation of the number of factors from using $\tilde{k}_{ER}$ is accurate.

Table 9: Test results for mortality rates			
	$\tilde{k}_{ER}$	T_{uf}	Conclusion
Mortality Rates	1	17.21101	False

8 Conclusions and Discussion

This paper establishes asymptotic properties for spiked empirical eigenvalues derived from sample covariance matrices for high-dimensional random vectors with dependent sample observations. From a

[†]Note that the death rate of 90+ is recalculated as d_{90+}/E_{90+} .

Figure 27: Eigenvalues of the sample covariance matrix for log death rates

statistical point of view, it is expected that empirical eigenvalues are able to accurately infer unknown population eigenvalues from the population covariance matrix under consideration. Unfortunately, on high-dimensional data, our results show that empirical eigenvalues are heavily affected by the dependent structure among sample observations. In particular, the empirical eigenvalues can only reflect the sample dependent structure under the weak cross-sectional dependence scenario.

PCA is an important dimensional reduction technique in high-dimensional statistics, which relies on large empirical eigenvalues to infer spiked population covariance structure. The asymptotic results in this paper demonstrate that PCA may provide inaccurate information on high-dimensional data with dependent sample observations. A popular example is high-dimensional non-stationary time series. We propose a test statistic to distinguish a unit root structure from a factor model, which is used to describe spiked population covariance structure or relatively strong cross-sectional dependence.

Furthermore, two sets of popular real data are investigated. The first empirical application on an OECD HE dataset provides statistical justification for many of the econometric models on this dataset proposed in the econometric literature. The second empirical analysis on the U.S. mortality data proves suitability of the benchmark model proposed by [25].

Appendix: Proofs of Theoretical Results

In this appendix, we provide the proofs of all the theorems and propositions.

Proof of Theorem 1

We first prove the following proposition to ensure the convergence in probability:

Proposition 5. For $\Gamma \mathbf{X} \Omega \mathbf{X}^* \Gamma^*$ satisfying Assumptions 1 and 4, the i -th largest eigenvalue λ_i converge to μ_i in probability after rescaling. More precisely, we have uniformly for all $i = 1, \dots, K$

$$\frac{\lambda_i - \mu_i \frac{\text{tr} \Omega}{n}}{\mu_i \frac{\text{tr} \Omega}{n}} \rightarrow 0 \text{ in probability.} \quad (8.1)$$

Proof. Recall K_ε in Assumption 4. For any $\varepsilon > 0$, we write $\Lambda = \begin{pmatrix} \Lambda_{S,\varepsilon} & 0 \\ 0 & \Lambda_{P,\varepsilon} \end{pmatrix}$, where $\Lambda_{S,\varepsilon} = \text{diag}(\mu_1, \dots, \mu_{K_\varepsilon-1})$ and $\Lambda_{P,\varepsilon} = \text{diag}(\mu_{K_\varepsilon}, \dots, \mu_{p-L})$. Note that the non-zero eigenvalues of $\Gamma \mathbf{X} \Omega \mathbf{X}^* \Gamma^*$ are the same as those of $\mathbf{U} \mathbf{X} \Omega \mathbf{X}^* \mathbf{U}^* \Lambda$. By weyl's inequality, we have

$$\left| \sigma_i(\Lambda^{1/2} \mathbf{U} \mathbf{X} \Omega^{1/2}) - \sigma_i \left(\begin{pmatrix} \Lambda_{S,\varepsilon}^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{U} \mathbf{X} \Omega^{1/2} \right) \right| \leq \sigma_1 \left(\begin{pmatrix} 0 & 0 \\ 0 & \Lambda_{P,\varepsilon}^{1/2} \end{pmatrix} \mathbf{U} \mathbf{X} \Omega^{1/2} \right),$$

where $\sigma_i(\mathbf{A})$ is the i -th largest singular value of $\mathbf{A}$. Let $\mathbf{P}_\varepsilon = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_{P,\varepsilon}^{1/2} \end{pmatrix} \mathbf{U} \mathbf{X} \Omega^{1/2}$.

We find that $\|\mathbf{P}_\varepsilon \mathbf{P}_\varepsilon^*\|_2 \leq \text{tr}(\mathbf{P}_\varepsilon \mathbf{P}_\varepsilon^*)$ and $\text{tr}(\mathbf{P}_\varepsilon \mathbf{P}_\varepsilon^*) \leq 2 \sum_{j=K_\varepsilon}^{p-L} \mu_j \frac{\text{tr}(\Omega)}{n}$ with probability tending to 1. Then, we find that when (4.1) holds, $\frac{\|\mathbf{P}_\varepsilon \mathbf{P}_\varepsilon^*\|_2}{\mu_K \frac{\text{tr}(\Omega)}{n}} \leq \varepsilon$ with probability tending to 1.

Moreover, we define $\mathbf{S}_\varepsilon = \begin{pmatrix} \Lambda_{S,\varepsilon}^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{U} \mathbf{X} \Omega^{1/2}$. By Theorem 7.1 of [5], we can show that $\frac{\sigma_i(S_\varepsilon)^2}{\mu_i \frac{\text{tr} \Omega}{n}} - 1 = o_p(1)$.

Then, we can conclude that

$$\lim_{p,n \rightarrow \infty} P(|\frac{\lambda_i}{\mu_i \frac{\text{tr} \Omega}{n}} - 1| \leq \varepsilon) = 1 \quad (8.2)$$

for any $\varepsilon > 0$. (8.1) is proved. □

We now prove CLT for a fixed λ_i , $i \in \{1, \dots, K\}$. Define $\mathbf{V}_1 = \Lambda_S^{1/2} \mathbf{U}_1 \mathbf{X}$ and $\mathbf{V}_2 = \Lambda_P^{1/2} \mathbf{U}_2 \mathbf{X}$. By the definition of λ_i , it solves the equation

$$\det(\lambda_i \mathbf{I} - \Lambda^{1/2} \mathbf{U} \mathbf{X} \Omega \mathbf{X}^* \mathbf{U}^* \Lambda^{1/2}) = 0.$$

Together with Proposition 5, with probability tending to 1, it is equivalent to

$$\det(\lambda_i \mathbf{I} - \mathbf{V}_1 \Omega \mathbf{V}_1^* - \mathbf{V}_1 \Omega \mathbf{V}_2^* (\lambda_i \mathbf{I} - \mathbf{V}_2 \Omega \mathbf{V}_2^*)^{-1} \mathbf{V}_2 \Omega \mathbf{V}_1^*) = 0.$$

It is easy to see $\|(\lambda_i \mathbf{I} - \mathbf{V}_2 \Omega \mathbf{V}_2^*)^{-1}\|_2 = O\left(\frac{1}{\mu_K \frac{\text{tr} \Omega}{n}}\right)$. Then we consider $\Lambda_S^{-1/2} \mathbf{V}_1 \Omega \mathbf{V}_2^* \mathbf{V}_2 \Omega \mathbf{V}_1^* \Lambda_S^{-1/2}$. Note that

$$\begin{aligned} \text{tr}(\Lambda_S^{-1/2} \mathbf{V}_1 \Omega \mathbf{V}_2^* \mathbf{V}_2 \Omega \mathbf{V}_1^* \Lambda_S^{-1/2}) &= \text{tr}(\mathbf{U}_1 \mathbf{X} \Omega \mathbf{X}^* \mathbf{U}_2^* \Lambda_P \mathbf{U}_2 \mathbf{X} \Omega \mathbf{X}^* \mathbf{U}_1^*) \\ &= \sum_{j=1}^K (\mathbf{U}_1 \mathbf{X} \Omega \mathbf{X}^* \mathbf{U}_2^* \Lambda_P \mathbf{U}_2 \mathbf{X} \Omega \mathbf{X}^* \mathbf{U}_1^*)_{jj} = \sum_{j=1}^K h_j, \end{aligned} \quad (8.3)$$

where $h_j = (\mathbf{U}_1 \mathbf{X} \mathbf{\Omega} \mathbf{X}^* \mathbf{U}_2^* \mathbf{\Lambda}_P \mathbf{U}_2 \mathbf{X} \mathbf{\Omega} \mathbf{X}^* \mathbf{U}_1^*)_{jj}$. Define the elements of $\mathbf{u}_j$ as u_{jk} . Write

$$\begin{aligned}
Eh_j &= E(\mathbf{u}_j^* \mathbf{X} \mathbf{\Omega} \mathbf{X}^* \mathbf{U}_2^* \mathbf{\Lambda}_P \mathbf{U}_2 \mathbf{X} \mathbf{\Omega} \mathbf{X}^* \mathbf{u}_j) \\
&= E(\mathbf{u}_j^* \mathbf{X} \mathbf{\Omega} \mathbf{X}^* \mathbf{\Sigma}_2 \mathbf{X} \mathbf{\Omega} \mathbf{X}^* \mathbf{u}_j) \\
&= \sum_{i_1, i_3=1}^T \sum_{j_1, j_3=1}^n u_{j, i_1} E(X_{i_1, j_1} \Omega_{j_1, j_1} X_{i_1, j_1} \Sigma_{i_1, i_3} X_{i_3, j_3} \Omega_{j_3, j_3} X_{i_3, j_3}) u_{j, i_3} \\
&\quad + \sum_{i_1, i_2=1}^T \sum_{j_1, j_2=1}^n u_{j, i_1} E(X_{i_1, j_1} \Omega_{j_1, j_2} X_{i_2, j_2} \Sigma_{i_2, i_1} X_{i_1, j_1} \Omega_{j_1, j_2} X_{i_2, j_2}) u_{j, i_2} \\
&\quad + \sum_{i_1, i_2=1}^T \sum_{j_1, j_2=1}^n u_{j, i_1} E(X_{i_1, j_1} \Omega_{j_1, j_2} X_{i_2, j_2} \Sigma_{i_2, i_2} X_{i_2, j_2} \Omega_{j_2, j_1} X_{i_1, j_1}) u_{j, i_1} \\
&\quad - 2 \sum_{i_1=1}^T \sum_{j_1=1}^n u_{j, i_1} E(X_{i_1, j_1} \Omega_{j_1, j_1} X_{i_1, j_1} \Sigma_{i_1, i_1} X_{i_1, j_1} \Omega_{j_1, j_1} X_{i_1, j_1}) u_{j, i_1} \\
&= I + II + III - IV.
\end{aligned} \tag{8.4}$$

Here Σ_{i_1, i_3} is the elements of $\mathbf{\Sigma}_2$. Note that $\mathbf{u}_j^* \mathbf{\Sigma}_2 \mathbf{u}_j = 0$.

$$\begin{aligned}
I &= \sum_{i_1, i_3=1}^T \sum_{j_1, j_3=1}^n u_{j, i_1} E(X_{i_1, j_1} \Omega_{j_1, j_1} X_{i_1, j_1} \Sigma_{i_1, i_3} X_{i_3, j_3} \Omega_{j_3, j_3} X_{i_3, j_3}) u_{j, i_3} \\
&= \sum_{i_1, i_3=1}^T \sum_{j_1, j_3=1}^n u_{j, i_1} \Sigma_{i_1, i_3} u_{j, i_3} \Omega_{j_1, j_1} \Omega_{j_3, j_3} E(X_{i_1, j_1}^2 X_{i_3, j_3}^2) \\
&= \sum_{i_1, i_3=1}^T \sum_{j_1, j_3=1}^n u_{j, i_1} \Sigma_{i_1, i_3} u_{j, i_3} \Omega_{j_1, j_1} \Omega_{j_3, j_3} E(X_{i_1, j_1}^2) E(X_{i_3, j_3}^2) \\
&\quad + \sum_{i_1=1}^T \sum_{j_1=1}^n u_{j, i_1} \Sigma_{i_1, i_1} u_{j, i_1} \Omega_{j_1, j_1} \Omega_{j_1, j_1} \{E(X_{i_1, j_1}^4) - E(X_{i_1, j_1}^2) E(X_{i_1, j_1}^2)\} \\
&\leq \frac{1}{n^2} \mathbf{u}_j^T \mathbf{\Sigma}_2 \mathbf{u}_j (\text{tr} \mathbf{\Omega})^2 + \frac{1}{n^2} \text{tr}(\mathbf{\Omega}^2) \text{tr}(\mathbf{\Sigma}_2) (\gamma_4 - 1) = O\left(\frac{\text{tr}(\mathbf{\Omega}^2) \text{tr}(\mathbf{\Sigma}_2)}{n^2}\right).
\end{aligned} \tag{8.5}$$

Similarly, noting that $\Sigma_{i_2, i_1} = \Sigma_{i_1, i_2}$ and $\Omega_{i_2, i_1} = \Omega_{i_1, i_2}$, we have

$$\begin{aligned}
II &= \sum_{i_1, i_2=1}^T \sum_{j_1, j_2=1}^n u_{j, i_1} E(X_{i_1, j_1} \Omega_{j_1, j_2} X_{i_2, j_2} \Sigma_{i_2, i_1} X_{i_1, j_1} \Omega_{j_1, j_2} X_{i_2, j_2}) u_{j, i_2} \\
&= \sum_{i_1, i_2=1}^T \sum_{j_1, j_2=1}^n u_{j, i_1} \Sigma_{i_1, i_2} u_{j, i_2} \Omega_{j_1, j_2}^2 E(X_{i_1, j_1}^2 X_{i_2, j_2}^2) = O\left(\frac{\text{tr}(\mathbf{\Omega}^2) \text{tr}(\mathbf{\Sigma}_2)}{n^2}\right),
\end{aligned} \tag{8.6}$$

$$\begin{aligned}
III &= \sum_{i_1, i_2=1}^T \sum_{j_1, j_2=1}^n u_{j, i_1} E(X_{i_1, j_1} \Omega_{j_1, j_2} X_{i_2, j_2} \Sigma_{i_2, i_2} X_{i_2, j_2} \Omega_{j_2, j_1} X_{i_1, j_1}) u_{j, i_1} \\
&= \sum_{i_1, i_2=1}^T \sum_{j_1, j_2=1}^n u_{j, i_1}^2 \Sigma_{i_2, i_2} \Omega_{j_1, j_2}^2 E(X_{i_1, j_1}^2 X_{i_2, j_2}^2) = O\left(\frac{\text{tr}(\mathbf{\Omega}^2) \text{tr}(\mathbf{\Sigma}_2)}{n^2}\right)
\end{aligned} \tag{8.7}$$

and

$$\begin{aligned}
IV &= 2 \sum_{i_1=1}^T \sum_{j_1=1}^n u_{j, i_1} E(X_{i_1, j_1} \Omega_{j_1, j_1} X_{i_1, j_1} \Sigma_{i_1, i_1} X_{i_1, j_1} \Omega_{j_1, j_1} X_{i_1, j_1}) u_{j, i_1} \\
&= 2 \sum_{i_1=1}^T \sum_{j_1=1}^n u_{j, i_1}^2 \Omega_{j_1, j_1}^2 \Sigma_{i_1, i_1} E(X_{i_1, j_1}^4) = O\left(\frac{\text{tr}(\mathbf{\Omega}^2) \text{tr}(\mathbf{\Sigma}_2)}{n^2}\right).
\end{aligned} \tag{8.8}$$

From (8.3)-(8.8) and h_j is non-negative, we conclude that

$$\text{tr}(\Lambda_S^{-1/2} \mathbf{V}_1 \Omega \mathbf{V}_2^* \mathbf{V}_2 \Omega \mathbf{V}_1^* \Lambda_S^{-1/2}) = O_P \left(\frac{\text{tr}(\Omega^2) \text{tr}(\Sigma_2)}{n^2} \right). \quad (8.9)$$

This, together with Assumptions 3-4, implies that

$$\|\sqrt{n} \Lambda_S^{-1/2} \mathbf{V}_1 \Omega \mathbf{V}_2^* (\lambda_i \mathbf{I} - \mathbf{V}_2 \Omega \mathbf{V}_2^*)^{-1} \mathbf{V}_2 \Omega \mathbf{V}_1^* \Lambda_S^{-1/2}\|_2 = O_P \left(\frac{\text{tr}(\Omega^2)}{\sqrt{n} \text{tr}(\Omega)} \right) = o_P \left(\sqrt{\frac{\text{tr}(\Omega^2)}{n}} \right). \quad (8.10)$$

Note that the variance of every elements in $\mathbf{R}_i$ is $\frac{\text{tr}(\Omega^2)}{n}$. (8.10) implies that the part can be ignored in CLT.

Define

$$Q = \frac{n}{\sqrt{\text{tr}(\Omega^2)}} \Lambda_S^{-1/2} \left(V_1 \Omega V_1^* - \text{diag} \left\{ \mu_1 \frac{\text{tr} \Omega}{n}, \dots, \mu_K \frac{\text{tr} \Omega}{n} \right\} \right) \Lambda_S^{-1/2} \quad (8.11)$$

and the $n_i \times n_i$ matrix R which has the element

$$R_{fg} = Q_{m_i+f, m_i+g}. \quad (8.12)$$

By Theorem 7.2 of [5] it is not difficult to conclude that

$$R \xrightarrow{d} \mathbf{R}_i. \quad (8.13)$$

Similarly, we can verify the order of the other elements in H and conclude the asymptotic independence between λ_{m_i+f} and λ_{m_j+g} .

Proof of Proposition 1

Proof of Proposition 1. Assumption 4 implies that for $\forall \varepsilon > 0$, there is K_ε , independent of n and T , such that

$$\frac{\mu_1}{\text{tr}(\Gamma \Gamma^*)} \geq \frac{\mu_1}{K_\varepsilon \mu_1 + \frac{\varepsilon}{2} \mu_1} = \frac{2}{\varepsilon + 2K_\varepsilon}. \quad (8.14)$$

Then there exists a positive constant c so that

$$\liminf_{T \rightarrow \infty} \frac{\mu_1}{\text{tr}(\Gamma \Gamma^*)} > c. \quad (8.15)$$

Similarly,

$$\lim_{n, T \rightarrow \infty} P \left(\frac{\lambda_1}{\text{tr}(\Gamma \mathbf{X} \Omega \mathbf{X}^* \Gamma^*)} > \frac{\lambda_1}{2 \text{tr}(\Gamma \Gamma^*) \frac{\text{tr}(\Omega)}{n}} \right) = 1, \quad (8.16)$$

$$\lim_{n, T \rightarrow \infty} P \left(\frac{\lambda_1}{\text{tr}(\Gamma \Gamma^*) \frac{\text{tr}(\Omega)}{n}} > \frac{\mu_1}{2 \text{tr}(\Gamma \Gamma^*)} \right) = 1 \quad (8.17)$$

and (8.15) implies that there exists a positive constant c such that

$$\lim_{T \rightarrow \infty} P \left(\frac{\lambda_1}{\text{tr}(\Gamma \mathbf{X} \Omega \mathbf{X}^* \Gamma^*)} > c \right) = 1. \quad (8.18)$$

$$\lim_{n, T \rightarrow \infty} \left(\frac{E \lambda_i}{E \lambda_{i+1}} - \frac{\mu_i}{\mu_{i+1}} \right) = 0 \quad (8.19)$$

and the standard deviation of λ_i has the smaller order than $E \lambda_i$. This implies (4.7). □

Proof of Theorem 2

We need to give a definition and the following two lemmas.

Definition 1. Define

$$\tilde{\mu}_i = \bar{\mu}_i \left(a_0 + 2 \sum_{j=1}^{T-1} a_j (-1)^j \cos \left(\frac{j(T-i)\pi}{T} \right) \right), \quad (8.20)$$

where $a_s = \sum_{i=0}^{L-s} b_i b_{i+s}$.

Lemma 1. Under Assumption 7 and Definition 1, for any fixed i ,

$$\frac{\mu_i - \tilde{\mu}_i}{\tilde{\mu}_i} = O\left(\frac{1}{T}\right) \quad (8.21)$$

and

$$\frac{\mu_i}{\mu_{i+1}} = \frac{1 + \cos\left(\frac{(T-i-1)\pi}{T}\right)}{1 + \cos\left(\frac{(T-i)\pi}{T}\right)} + O\left(\frac{1}{T}\right) = \frac{\bar{\mu}_i}{\bar{\mu}_{i+1}} + O\left(\frac{1}{T}\right). \quad (8.22)$$

Moreover, for any fixed K , when T is large enough, $\mu_1 > \mu_2 > \dots > \mu_K$ satisfy Assumptions 2 and 4.

Lemma 2. Under Assumptions 5-7, for any fixed i ,

$$\lim_{n \rightarrow \infty} \frac{n^2}{\text{tr}(\mathbf{\Omega}^2)} \times \text{Var}(\mathbf{u}_i^* \mathbf{X} \mathbf{\Omega} \mathbf{X}^* \mathbf{u}_i) = 2. \quad (8.23)$$

The ideas of the proofs are as follows. At first we study the eigenvalues and eigenvectors of $\mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C}$. Then we establish a relation between $\mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C}$ and $\mathbf{W} \mathbf{W}^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C}$. Note that $\mathbf{W} \mathbf{W}^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C}$ has the same non-zero eigenvalues as $\mathbf{H} \mathbf{C} \mathbf{W} \mathbf{W}^* \mathbf{C}^* \mathbf{H}^*$.

Lemma 3. $\bar{\mu}_1 > \bar{\mu}_2 > \dots > \bar{\mu}_{T-1}$ are the non-zero eigenvalues of $\mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C}$.

Lemma 4 below specifies the eigenvectors of $\mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C}$.

Lemma 4. Let $\check{\mathbf{x}}_{\mathbf{k}} = (0, \check{x}_{k,1}, \dots, \check{x}_{k,T-1})^*$ be a $T \times 1$ vector with

$$\check{x}_{k,i} = (-1)^{T-i} \sin\left(\frac{(T-i)(T-k)\pi}{T}\right), \quad -T \leq i \leq 2T. \quad (8.24)$$

Then $\{\check{\mathbf{x}}_{\mathbf{k}}, 1 \leq k \leq T-1\}$ are orthogonal and satisfy for any k

$$\mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \check{\mathbf{x}}_{\mathbf{k}} = \bar{\mu}_k \check{\mathbf{x}}_{\mathbf{k}}. \quad (8.25)$$

Lemmas 3 and 4 are proved in [40]. Now we prove Lemma 1.

Proof of Lemma 1. Note that

$$\check{x}_{k,i} = (-1)^{T-i} \sin\left(\frac{(T-i)(T-k)\pi}{T}\right) = (-1)^{2T-i-k+1} \sin\left(\frac{i(T-k)\pi}{T}\right).$$

Then

$$\check{x}_{k,i} = (-1)^{k-1} \check{x}_{k,T-i} = (-1)^k \check{x}_{k,T+i} = -\check{x}_{k,2T-i} = -\check{x}_{k,-i}.$$

$$\begin{aligned}
\ddot{x}_{k,s-j} + \ddot{x}_{k,s+j} &= (-1)^{T-s-j} \left(\sin \left(\frac{(T-s+j)(T-k)\pi}{T} \right) + \sin \left(\frac{(T-s-j)(T-k)\pi}{T} \right) \right) \\
&= 2(-1)^{T-s-j} \sin \left(\frac{(T-s)(T-k)\pi}{T} \right) \cos \left(\frac{j(T-k)\pi}{T} \right) \\
&= 2(-1)^j \ddot{x}_{k,s} \cos \left(\frac{j(T-k)\pi}{T} \right).
\end{aligned} \tag{8.26}$$

Now we define $\mathbf{A} = \mathbf{W}\mathbf{W}^*$ and one can verify that the entries of $\mathbf{A}$ satisfy

$$A_{ij} = a_{|i-j|}.$$

We also define another matrix $\mathbf{A}_y$ with the entries

$$A_{y,ij} = \begin{cases} -a_{i+j-2} & i+j \leq T+1, \\ -a_{2T-j-i+2} & T+3 \leq i+j, \\ 0 & \text{otherwise.} \end{cases} \tag{8.27}$$

Then, one can verify that

$$(\mathbf{A} + \mathbf{A}_y)\mathbf{C}^*\mathbf{H}^*\mathbf{H}\mathbf{C}\ddot{\mathbf{x}}_k = \bar{\mu}_k \left(a_0 + 2 \sum_{j=1}^{T-1} a_j (-1)^j \cos \left(\frac{j(T-k)\pi}{T} \right) \right) \ddot{\mathbf{x}}_k$$

by (8.26). From Assumption 7 and Definition 1, we can conclude that for any fixed k , the k th largest eigenvalue of $(\mathbf{A} + \mathbf{A}_y)\mathbf{C}^*\mathbf{H}^*\mathbf{H}\mathbf{C}$ is $\tilde{\mu}_k$. Note that, $\mathbf{H}\mathbf{C}\mathbf{A}_y\mathbf{C}^*\mathbf{H}^*$ is symmetry and

$$|\mathbf{H}\mathbf{C}\mathbf{A}_y\mathbf{C}^*\mathbf{H}^*|_2 \leq T \max_{i,j} \{(\mathbf{H}\mathbf{C}\mathbf{A}_y\mathbf{C}^*\mathbf{H}^*)_{i,j}\} \leq 2T \sum_{i=0}^{T-1} i|a_i| = O(T).$$

Then we can conclude that for any fixed k ,

$$\mu_k = \tilde{\mu}_k + O(T).$$

Note that $\bar{\mu}_k$ has a order of T^2 with any fixed k and so we prove (8.21) and (8.22).

Since $\lim_{T \rightarrow \infty} \frac{\bar{\mu}_i}{\bar{\mu}_{i+1}} = \frac{(i+1)^2}{i^2}$ for any fixed i . $\mu_1 > \mu_2 > \dots > \mu_K$ satisfy Assumption 2. Moreover, one can verify Assumption 4 by calculating $\text{tr}(\mathbf{H}\mathbf{C}\mathbf{W}\mathbf{W}^*\mathbf{C}^*\mathbf{H}^*)$ and $\sum_{i=1}^{\infty} \tilde{\mu}_i$. □

Proof of Lemma 2. Define the elements of $\mathbf{u}_i$ as u_{it} . Then

$$\mathbf{u}_i^* \mathbf{X} \mathbf{\Omega} \mathbf{X}^* \mathbf{u}_i = \sum_{t_1=1-L}^T \sum_{t_2=1-L}^T \sum_{s_1=1}^n \sum_{s_2=1}^n u_{it_1} x_{s_1 t_1} \Omega_{s_1 s_2} x_{s_2 t_2} u_{it_2}. \tag{8.28}$$

It follows that

$$\begin{aligned}
\text{var}(\mathbf{u}_i^* \mathbf{X} \mathbf{\Omega} \mathbf{X}^* \mathbf{u}_i) &= \sum_{t_1=1-L}^T \sum_{t_2=1-L}^T \sum_{s_1=1}^n \sum_{s_2=1}^n u_{it_1}^2 \Omega_{s_1 s_2}^2 u_{it_2}^2 E(x_{s_1 t_1}^2) E(x_{s_2 t_2}^2) \\
&+ \sum_{t_1=1-L}^T \sum_{t_2=1-L}^T \sum_{s_1=1}^n \sum_{s_2=1}^n u_{it_1}^2 \Omega_{s_1 s_2} \Omega_{s_2 s_1} u_{it_2}^2 E(x_{s_1 t_1}^2) E(x_{s_2 t_2}^2) \\
&+ \sum_{t=1-L}^T \sum_{s=1}^n u_{it}^4 \Omega_{ss}^2 (E(x_{st}^4) - E(x_{st}^2) E(x_{st}^2)) \\
&= \frac{2}{n^2} \text{tr}(\mathbf{\Omega}^2) + \sum_{t=1-L}^T \sum_{s=1}^n u_{it}^4 \Omega_{ss}^2 \frac{\gamma_4 - 1}{n^2}.
\end{aligned} \tag{8.29}$$

So we only need to prove $\sum_{t=1-L}^T u_{it}^4 = o(1)$. Note that $\sum_{t=1-L}^T u_{it}^2 = 1$, we only need to prove $\max_{1-L \leq t \leq T} \{ |u_{it}| \} = o(1)$.

Now we define $\mathbf{f}_k$ to satisfy that $\|\mathbf{f}_k\|_2 = 1$ and

$$\mathbf{A}\mathbf{C}^*\mathbf{H}^*\mathbf{H}\mathbf{C}\mathbf{f}_k = \mu_k \mathbf{f}_k. \quad (8.30)$$

We then need to verify that

$$\mathbf{H}\mathbf{C}\mathbf{A}\mathbf{C}^*\mathbf{H}^* \frac{\mathbf{H}\mathbf{C}\mathbf{f}_k}{\|\mathbf{H}\mathbf{C}\mathbf{f}_k\|_2} = \mu_k \frac{\mathbf{H}\mathbf{C}\mathbf{f}_k}{\|\mathbf{H}\mathbf{C}\mathbf{f}_k\|_2} \quad (8.31)$$

and

$$\mathbf{u}_k = \frac{\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{W}}{\|\mathbf{H}\mathbf{C}\mathbf{f}_k\|_2 \sqrt{\mu_k}}. \quad (8.32)$$

We also define $\ddot{\mathbf{y}}_k = \frac{\ddot{\mathbf{x}}_k}{\|\ddot{\mathbf{x}}_k\|_2}$ and rewrite

$$\mathbf{f}_k = \sum_{j=1}^T a_{kj} \ddot{\mathbf{y}}_k \quad (8.33)$$

where

$$\sum_{j=1}^T a_{kj}^2 = 1. \quad (8.34)$$

Then

$$\begin{aligned} \mu_k &= \frac{\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{A} \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k}{\|\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k\|_2} \\ &= \frac{\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} (\mathbf{A} + \mathbf{A}_y - \mathbf{A}_y) \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k}{\|\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k\|_2^2}. \end{aligned} \quad (8.35)$$

It follows that

$$\begin{aligned} &\frac{|\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} (\mathbf{A} + \mathbf{A}_y) \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k| - |\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{A}_y \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k|}{\|\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k\|_2^2} \leq \mu_k \\ &\leq \frac{|\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} (\mathbf{A} + \mathbf{A}_y) \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k| + |\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{A}_y \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k|}{\|\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k\|_2^2}. \end{aligned} \quad (8.36)$$

Lemma 4 and (8.33) imply that

$$\|\mathbf{H}\mathbf{C}\mathbf{f}_k\|_2 = \sqrt{\sum_{j=1}^T a_{kj}^2 \bar{\mu}_j} \quad (8.37)$$

and

$$\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} (\mathbf{A} + \mathbf{A}_y) \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k = \sum_{j=1}^T a_{kj}^2 \tilde{\mu}_j \bar{\mu}_j. \quad (8.38)$$

Note that

$$\frac{|\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{A}_y \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k|}{\|\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{f}_k\|_2^2} \leq \|\mathbf{H} \mathbf{C} \mathbf{A}_y \mathbf{C}^* \mathbf{H}^*\|_2 = O(T) = O\left(\frac{\mu_k}{T}\right). \quad (8.39)$$

Then (8.36)-(8.39) imply that

$$\mu_k = \frac{\sum_{j=1}^T a_{kj}^2 \tilde{\mu}_j \bar{\mu}_j}{\sum_{j=1}^T a_{kj}^2 \bar{\mu}_j} + O\left(\frac{\mu_k}{T}\right). \quad (8.40)$$

Note that for any fixed i ,

$$\lim_{T \rightarrow \infty} \frac{\bar{\mu}_i}{\bar{\mu}_{i+1}} = \lim_{T \rightarrow \infty} \frac{\tilde{\mu}_i}{\tilde{\mu}_{i+1}} = \lim_{T \rightarrow \infty} \frac{\mu_i}{\mu_{i+1}} = \frac{(i+1)^2}{i^2}.$$

Then one can verify that for any fixed k ,

$$a_{kk} = 1 + O\left(\frac{1}{T}\right). \quad (8.41)$$

By (8.32)-(8.33) and (8.37), we can conclude that

$$\mathbf{u}_k = \frac{\mathbf{f}_k^* \mathbf{C}^* \mathbf{H}^* \mathbf{H} \mathbf{C} \mathbf{W}}{\|\mathbf{H} \mathbf{C} \mathbf{f}_k\|_2 \sqrt{\mu_k}} = \frac{\sum_{j=1}^T a_{kj} \bar{\mu}_k \ddot{\mathbf{y}}_k \mathbf{W}}{\sqrt{\sum_{j=1}^T a_{kj}^2 \bar{\mu}_j \mu_k}}.$$

This, together with (8.41), implies that $\max_{1-L \leq t \leq T} \{|u_{kt}|\} = o(1)$. $\square$

Then, along with Theorem 1, Lemmas 1 and 2 conclude Theorem 2.

Proof of Propositions 3

Note that, when i is fixed,

$$\lim_{T \rightarrow \infty} \frac{\bar{\mu}_i}{\bar{\mu}_{i+1}} = \lim_{T \rightarrow \infty} \frac{1 + \cos\left(\frac{(T-i-1)\pi}{T}\right)}{1 + \cos\left(\frac{(T-i)\pi}{T}\right)} = \frac{(i+1)^2}{i^2}. \quad (8.42)$$

This, together with Theorem 2 and Lemma 1, implies (5.14).

Moreover, Lemma 1 implies

$$T_{uf} = \frac{\lambda_1 \frac{\bar{\mu}_2}{\bar{\mu}_1} - \lambda_2}{\lambda_2 - \lambda_3 \frac{\bar{\mu}_2}{\bar{\mu}_3}} = \frac{\lambda_1 \frac{\mu_2}{\mu_1} - \lambda_2 + \lambda_1 O(\frac{1}{T})}{\lambda_2 - \lambda_3 \frac{\mu_2}{\mu_3} + \lambda_3 O(\frac{1}{T})}. \quad (8.43)$$

This, together with Theorem 2 and (5.15), implies Proposition 3.

Proof of Propositions 4

Under Assumption 8, the largest eigenvalue of the single factor model has a higher order than others and the difference between the second and the third largest eigenvalue is bounded. Then we have Propositions 4.

References

- [1] ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34**, 122-148.
- [2] AHN, S. C. and HORENSTEN, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203-1227.
- [3] BAI, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *J. Econometr.* **122**, 137-183.
- [4] BAI, Z. D. and SILVERSTEIN, J. W. (2006). Spectral analysis of large-dimensional random matrices, **2nd Edition**. *Springer, New York*.

- [5] BAI, Z. D. and YAO, J. F. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. H. Poincaré* **44**, 447-474.
- [6] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for non-null complex sample covariance matrices. *Ann. Probab.* **33**, 1643-1697.
- [7] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97**, 1382-1408.
- [8] BALTAGI, B. H. and MOSCONE, F. (2010). Healthcare expenditure and income in the OECD reconsidered: evidence from panel data. *Economic Modelling* **27**, 804-811.
- [9] BAO, Z. G., PAN, G. M. and ZHOU, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43**, 382-421.
- [10] CAI, T., HAN, X. and PAN, G. M. (2018). Limiting laws for divergent spiked eigenvalues and largest non-spiked eigenvalue of sample covariance matrices. *Forthcoming in Ann. Statist.*
- [11] CHAN, N.H. and WEI, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367-401.
- [12] CHEN, B. B. and PAN, G. M. (2012). Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero. *Bernoulli* **18**, 1405-1420.
- [13] DICKEY, D. A. and FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74**, 423-431.
- [14] EL KAROUI, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35**, 663-714.
- [15] FAN, J. Q., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Statist. Soc. B* **75**, 603-680.
- [16] FAN, J. Q., WANG, D., WANG, K. and ZHU, Z. (2017). Distributed estimation of principal eigenspaces. *Forthcoming in Ann. Statist.*
- [17] FISHER, R. A. (1937). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugenics.* **9**, 238-249.
- [18] GAO, J. T., XIA, K. and ZHU, H. J. (2019). Heterogeneous panel data models with cross-sectional dependence. Available at <https://ideas.repec.org/p/msh/ebswps/2017-16.html>. *Forthcoming in J. Econometr.*
- [19] HAN, F. and LIU, H. (2018). ECA: High-dimensional elliptical component analysis in non-Gaussian distributions. *J. Amer. Statist. Assoc.* **113**, 252-268.
- [20] HAN, X., PAN, G. M. and ZHANG, B. (2016). The Tracy-Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.* **44**, 1564-1592.
- [21] HSU, P. L. (1939). On the distribution of roots of certain determinant equations. *Ann. Eugenics.* **9**, 250-258.
- [22] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.* **29**, 295-327.
- [23] JOHNSTONE, I. M. (2007). High-dimensional statistical inference and random matrices. *Proceedings of the International Congress of Mathematicians Madrid* (available at DOI: 10.4171/022-1/13), 307-333.

- [24] KNOWLES, A. and YIN, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Rel.* **169**, 257.
- [25] LEE, R. and CARTER, L. (1992). Modeling and forecasting U.S. mortality. *J. Amer. Statist. Assoc.* **419**, 659-671.
- [26] LIU, H. Y., AUE, A. and PAUL, D. (2015) On the Marcenko-Pastur law for linear time series. *Ann. Statist.* **43**, 675-712.
- [27] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution for some sets of random matrices. *Math. USSR-Sb.* **1**, 457-483.
- [28] ONATSKI, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica* **77**, 1447-1479.
- [29] ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. and Statist.* **92**, 1004-1016.
- [30] PAN, G. M., GAO, J. and YANG, Y. R. (2014). Testing independence among a large number of high-dimensional random vectors. *J. Amer. Statist. Assoc.* **109**, 600-612.
- [31] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica.* **17**, 1617-1642.
- [32] PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: a review. *J. Statist. Plan. & Infer.* **150**, 1-29.
- [33] ROY, S. N. (1939). P -statistics and some generalizations in analysis of variance appropriate to multivariate problems. *Sankhyā.* **4**, 381-396.
- [34] SOSHIKOV, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108**, 1033-1056.
- [35] TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159**, 151-174.
- [36] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177**, 727-754.
- [37] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high-dimensional spiked covariance. *Ann. Statist.* **45**, 1342-1374.
- [38] YAO, J. F., ZHENG, S. R. and BAI, Z. D. (2015). Large sample covariance matrices and high-dimensional data analysis. *Cambridge University Press*.
- [39] ZHANG, B., PAN, G. M. and GAO, J. (2018). CLT for largest eigenvalues and unit root tests for high-dimensional nonstationary time series. *Ann. Statist.* **46**, 2186-2215.
- [40] ZHANG, B., PAN, G. M., and GAO, J. (2018). Supplement to "CLT for largest eigenvalues and unit root tests for high-dimensional nonstationary time series". Available at [doi:10.1214/17-AOS1616SUPP](https://doi.org/10.1214/17-AOS1616SUPP).
- [41] ZHANG, B., GAO, J., and PAN, G. M. (2019). A near unit root test for high-dimensional nonstationary time series. Available at <https://ideas.repec.org/p/msh/ebswps/2019-10.html>.
- [42] ZHANG, L. (2006). Spectral analysis of large dimensional random matrices. *Ph.D. Thesis, National University of Singapore, Singapore*.