Submitted to the Annals of Statistics

CLT FOR LARGEST EIGENVALUES AND UNIT ROOT TESTING FOR HIGH–DIMENSIONAL NONSTATIONARY TIME SERIES

By Bo Zhang, Guangming Pan* and Jiti Gao†

Let $\{Z_{ij}\}$ be independent and identically distributed (i.i.d.) random variables with $EZ_{ij} = 0$, $E|Z_{ij}|^2 = 1$ and $E|Z_{ij}|^4 < \infty$. Define linear processes $Y_{tj} = \sum_{k=0}^{\infty} b_k Z_{t-k,j}$ with $\sum_{i=0}^{\infty} |b_i| < \infty$. Consider a p–dimensional time series model of the form: $\mathbf{x}_t = \mathbf{\Pi}\mathbf{x}_{t-1} + \Sigma^{1/2}\mathbf{y}_t,\ 1 \le t \le T$ with $\mathbf{y}_t = (Y_{t1}, \cdots, Y_{tp})'$ and $\Sigma^{1/2}$ be the square root of a symmetric positive definite matrix. Let $\mathbf{B} = (1/p)\mathbf{X}\mathbf{X}^*$ with $\mathbf{X} = (\mathbf{x_1}, \cdots, \mathbf{x_T})'$ and X^* be the conjugate transpose. This paper establishes both the convergence in probability and the asymptotic joint distribution of the first k largest eigenvalues of $\mathbf{B}$ when $\mathbf{x}_t$ is nonstationary. As an application, one new unit root test for possible nonstationarity of high–dimensional time series is proposed and then studied both theoretically and numerically.

1. Introduction. There have been an increasing interest and significant developments on the theory and methodologies for handling high-dimensional data in recent years. Understanding high-dimensional sample covariance matrices, including its eigenvalues and eigenvectors, has proved to be extremely useful for such developments. Indeed, random matrix theory has provided useful estimation and testing procedures for high-dimensional data analysis. Recent discussions on this topic can be found in Johnstone [18], Paul and Aue [26] and Yao, Zheng and Bai [35].

Research towards understanding the eigenvalues of sample covariance matrices dates back to as early as the studies of Fisher [14], Hsu [15] and Roy [31], and has become increasingly active since the publication of the celebrated work of Marcenko and Pastur [22], in which the authors established a limiting spectral distribution (MP type distribution) for a sample covariance matrix for the case where p and T are comparable. More recent research has been devoted to establishing asymptotic properties for the eigenvalues and eigenvectors of high-dimensional sample covariance matrices.

*Partially supported by a MOE Tier 2 grant 2014-T2-2-060 and by a MOE Tier 1 Grant RG25/14 at the Nanyang Technological University, Singapore.

†Partially supported by the Australian Research Council Discovery Grants Program under Grant numbers: DP150101012 & DP170104421.

MSC 2010 subject classifications: Primary 60B20, 34K25; secondary 60F05, 62H10

Keywords and phrases: Asymptotic normality, Largest eigenvalue, Linear process, Unit root test

There are currently two main lines of research about asymptotic distributions of the largest eigenvalues of high–dimensional random matrices. The first line of research is concerned with the Tracy-Widom law of the largest eigenvalues of random matrices. It is well known that limiting distributions of the largest eigenvalues of high-dimensional random matrices, such as Wigner matrices, follow the Tracy-Widom law, which was originally discovered by Tracy and Widom in [33] and [34] for Gaussian Wigner ensembles. The largest eigenvalue of the Wishart matrix was investigated in Johnstone [17]. Several advancements for general sample covariance matrices have also been made, and we refer to [5] and [13] among others.

Empirical data from wireless communication, finance and speech recognition often suggest that some extreme eigenvalues of sample covariance matrices are well separated from the rest. This intrigues the second line of research about the spiked eigenvalues, which was first proposed in Johnstone [17]. The recent literature focuses on studying the behaviour of these spiked eigenvalues. For instance, the CLTs of the largest eigenvalues of complex Gaussian sample covariance matrices with a spiked population were investigated in Baik et al. [3], which also reported an interesting phase transition phenomenon. Baik and Silverstein [4] further considered almost sure limits of the extreme sample eigenvalues of the general spiked population. Paul [25] established a CLT for the spiked eigenvalues under the Gaussian population and the population spikes being simple. The fluctuation of the extreme sample eigenvalues of the general spiked population with arbitrary multiplicity numbers was further reported in Bai and Yao [2].

Most of the above existing studies rely on the assumption that the observations of high dimensional data are independent, although dimensional correlation structure can be allowed. Observations of high dimensional data in economics and finance, for example, are often highly dependent across time. In view of this, Zhang [36] investigated the empirical spectral distribution (ESD) of the sample covariance for the case where the data matrices are of the form $\mathbf{A}_1\mathbf{Z}\mathbf{A}_2$, where $\mathbf{A}_1$ and $\mathbf{A}_2$ are positive semidefinite matrices and $\mathbf{Z}$ has independent entries satisfying some moment assumptions. This model is referred to as a separable covariance model and allows for some dependence among observations recorded over different time points. Liu, Aue and Paul [20] studied the ESD of sample covariance matrices and symmetrized sample autocovariance matrices constructed from a linear process. Note that their setting also accommodates dependence among observations due to the fact that linear processes are built from the same innovation vectors. However, the above two papers only considered the ESD.

To the best of our knowledge, there is no existing work available to deal with the largest eigenvalues of sample covariance matrices generated from high dimensional nonstationary time series data. The main difficulty is that the properties of the population covariance matrices of the non-stationary data are unknown yet (even through we may make some assumptions about the error process). This paper belongs to the second line of research about the spiked eigenvalues. The main contribution of this paper is to establish several joint asymptotic distributions for the first several largest eigenvalues of sample covariance matrices of high dimensional nonstationary time series data. An additional contribution of this paper is to propose two new unit root tests for testing nonstationarity of high dimensional dependent time series.

We conclude this section by giving its organization. Section 2 establishes an asymptotic distributional theory for the first several largest eigenvalues of the covariance matrix of a high dimensional dependent time series. Section 3 proposes a new unit root test that is devoted to testing nonstationarity for high dimensional dependent data. Section 4 evaluates both the size and power properties of the proposed tests. Section 5 gives some concluding remarks. Appendix A establishes some useful results for truncated versions of sample covariance matrices by truncating linear processes. Appendix B gives the full proofs of the main theorems in Section 3. The proofs of the results listed in Appendix A are given in Appendix C of a supplementary document. Appendix D of the supplementary document discusses some possible extensions of the main models to include a cointegrating structure and a deterministic trending component.

2. Asymptotic Theory.

This section first introduces some necessary assumptions before we establish new asymptotic properties for the largest eigenvalues of the covariance matrix of a vector of high dimensional time series.

2.1. *Matrix models.*

The paper is to consider high dimensional covariance matrices for nonstationary time series. Specifically, define the following linear processes:

$$Y_{tj} = \sum_{k=0}^{\infty} b_k Z_{t-k,j} \tag{2.1}$$

with $\sum_{i=0}^{\infty} |b_i| < \infty$. Suppose that $\mathbf{y}_t = (Y_{t1}, \cdots, Y_{tp})'$ is a p–dimensional time series, where $\{Z_{ij}\}$ are independent and identically distributed (i.i.d.) random variables with $EZ_{ij} = 0$, $E|Z_{ij}|^2 = 1$ and $E|Z_{ij}|^4 < \infty$. Consider a

p–dimensional time series model of the form:

$$\mathbf{x}_t = \mathbf{\Pi}\mathbf{x}_{t-1} + \mathbf{\Sigma}^{1/2}\mathbf{y}_t,\ 1 \le t \le T, \tag{2.2}$$

where the spectral norm of the coefficient matrix $\mathbf{\Pi}$ is bounded by one ($0 \le \|\mathbf{\Pi}\|_2 \le 1$).

Define $\bar{\mathbf{X}} = \left(\frac{\sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{x_t}}{\mathbf{T}}, \cdots, \frac{\sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{x_t}}{\mathbf{T}}\right)'$ as a $T \times p$ matrix. Introduce the non-centered and centered sample covariance matrices

$$\mathbf{B} = \frac{1}{p}\mathbf{X}\mathbf{X}^* \tag{2.3}$$

and

$$\bar{\mathbf{B}} = \frac{1}{p}(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^* \tag{2.4}$$

with $\mathbf{X} = (\mathbf{x_1}, \cdots, \mathbf{x_T})'$. Here we point out that when $\mathbf{\Pi} = \mathbf{0}$, $\mathbf{\Sigma}$ satisfies some conditions and Y_{tj}'s are i.i.d random variables, the Tracy-Widom distribution has been established for the large eigenvalue of $\mathbf{B}$ in [5]. Also, when $\mathbf{\Pi} = \mathbf{0}$, $\mathbf{\Sigma}$ is a block matrix with spiked eigenvalues and Y_{tj}'s are i.i.d random variables, an asymptotic distribution (Gaussian distribution under some conditions) for the largest eigenvalues of $\mathbf{B}$ has been discussed in [25] and [2]. It is not clear yet how the largest eigenvalues of $\mathbf{B}$ may behave when Y_{tj}'s have some dependence structure. One case is that $\mathbf{\Pi} = \mathbf{0}$, but $\mathbf{\Sigma}$ is involved in (2.1). When $\mathbf{\Pi} = \mathbf{I}$, (2.2) becomes nonstationary. The main motivation for considering such a model is the proposal of two unit root tests to be discussed in the next section.

This paper is to investigate the largest eigenvalues of $\mathbf{B}$ and $\bar{\mathbf{B}}$ for the cases where $\mathbf{\Pi} = \mathbf{I}$ or $\|\mathbf{\Pi}\|_2 = \varphi < 1$. Throughout the paper, we make the following assumptions about the coefficients b_i and $\mathbf{\Sigma}$:

(A1) $\sum_{i=0}^{\infty} i|b_i| < \infty$.
(A2) $\sum_{i=0}^{\infty} b_i = s \neq 0$.
(A3) There exist two positive constants M_0 and M_1 such that $\|\mathbf{\Sigma}\|_2 \le M_0$ and $\frac{tr(\mathbf{\Sigma})}{p} \ge M_1$.
(A4) Let $T \to \infty$ and $p \to \infty$ such that $\lim_{T,\ p\to\infty} \frac{\sqrt{p}}{T} = 0$.

Here $\|\cdot\|_2$ stands for either the spectral norm of a matrix or the Euclidean norm of a vector. The linear process includes both MA(q) and AR(1) models. Assumption A2 is easily satisfied. Note that we do not require p and T to be of the same order, which is being commonly used in the random matrix theory literature. Assumption A3 covers some commonly used $\mathbf{\Sigma}$.

For example one may verify that the identity matrix $\mathbf{I}$ and the Toeplitz matrices satisfy it. However, we point out that Assumption A3 rules out the case where cross-sectional dependence has a factor model structure, which leads to very large eigenvalues of $\mathbf{\Sigma}$. We also need to make some assumptions about Z_{ij} and $\mathbf{x_0}$.

(A5) $\{Z_{i,j}\}$ are i.i.d random variables with mean zero, variance one and bounded fourth moment. Let $\mathbf{z}_t = (Z_{t1}, \cdots, Z_{tp})'$, where t can be either positive or negative integer (for the purpose of introducing A7 below).

(A6) $E\|\mathbf{x_0}\|_2^2 = O(p)$.

(A7) $\mathbf{x_0} = \sum\limits_{k=0}^{\infty} \tilde{b}_k \mathbf{\Sigma_1}^{1/2} \mathbf{z_{-k}} + \tilde{b}_{-1} \mathbf{\Sigma_2}^{1/2} \tilde{\mathbf{z}} + \tilde{\mathbf{b}}_{-2}$, where $\|\mathbf{\Sigma_1}\|_2 \leq M_0$, $\|\mathbf{\Sigma_2}\|_2 \leq M_0$ and $\tilde{\mathbf{z}} = (\tilde{Z}_1, \cdots, \tilde{Z}_p)'$ is independent of $\mathbf{z_t}$ for any t, in which $\{\tilde{Z}_j\}$ are i.i.d random variables with mean zero, variance one and finite fourth moments. The coefficients satisfy $\sum\limits_{k=0}^{\infty} |\tilde{b}_k| + |\tilde{b}_{-1}| < \infty$ and $\|\tilde{\mathbf{b}}_{-2}\|^2 = O(p)$.

2.2. *Main results for non-centered sample covariance matrix* $\mathbf{B}$. To characterize the limits in probability for the eigenvalues of $\mathbf{B}$, define for $k = 1, \cdots, T$,

$$\lambda_k = \frac{1}{2(1+\cos\theta_k)} \quad \text{with} \quad \theta_k = \frac{2(T+1-k)\pi}{2T+1}, \tag{2.5}$$

and

$$\gamma_k = \lambda_k \left(a_0 + 2\sum_{j=1}^{\infty} a_j (-1)^j \cos(j\theta_k) \right), \tag{2.6}$$

where

$$a_i = \sum_{k=0}^{\infty} b_k b_{k+i}. \tag{2.7}$$

We first characterize the magnitude of λ_k and γ_k.

PROPOSITION 1. *Let Assumptions A1 and A2 hold. For any fixed constant $k \geq 1$, there is a constant c_k such that*

$$\lim_{T\to\infty} \frac{\gamma_k}{T^2} = c_k > 0 \tag{2.8}$$

and

$$\lim_{T\to\infty}\frac{\gamma_k}{\gamma_1}=\lim_{T\to\infty}\frac{\lambda_k}{\lambda_1}=\frac{1}{(2k-1)^2}. \tag{2.9}$$

We are now at a position to state the main results; their proofs are given in Appendix B. The first theorem develops an upper bound in probability for the spectral norm of $\mathbf{B}$ for the stationary case. The second theorem gives a limit in probability and a joint distribution for the first k largest eigenvalues of $\mathbf{B}$ for nonstationary data.

THEOREM 2.1. *Let Assumptions A1–A6 hold. When* $0 \le \|\mathbf{\Pi}\|_2 = \varphi < 1$, *we obtain*

$$\|\mathbf{B}\|_2 = O_p\left(\frac{\left(1+\sqrt{\frac{T}{p}}\right)^2}{(1-\varphi)^2}\right). \tag{2.10}$$

THEOREM 2.2. *Let Assumptions A1–A5 hold. Let* ρ_k *be the* k*th largest eigenvalue of* $\mathbf{B}$. *Let* $\mathbf{\Pi}=\mathbf{I}$ *and* k *is fixed.*

(1) If Assumptions A6 holds, we have

$$\frac{\rho_k-\gamma_k\frac{tr(\mathbf{\Sigma})}{p}}{\gamma_1}\xrightarrow{i.p.}0, \tag{2.11}$$

where i.p. means convergence in probability.

(2) If Assumptions A7 holds, the random vector

$$\frac{\sqrt{p}}{\gamma_1}\left(\rho_1-\gamma_1\frac{tr(\mathbf{\Sigma})}{p},\cdots,\rho_k-\gamma_k\frac{tr(\mathbf{\Sigma})}{p}\right)' \tag{2.12}$$

converges weakly to a zero-mean Gaussian vector $\mathbf{w}=(w_1,\cdots,w_k)'$ *with covariance function* $\mathrm{cov}(w_i,w_j)=0$ *for any* $i\neq j$ *and* $\mathrm{var}(w_i)=\frac{2\theta}{(2i-1)^4}$ *with* $\theta=\lim_{p\to\infty}\frac{tr(\mathbf{\Sigma}^2)}{p}$.

REMARK 1. *The result holds for the complex case as well. In fact when* $\mathbf{Z}$ *is complex, set*

$$Re(Z_{jk})=Z_{ij}^R,\quad and\quad Im(Z_{jk})=Z_{ij}^I. \tag{2.13}$$

Let Z_{ij}^R *and* Z_{ij}^I *be independent. Then* $\frac{\sqrt{p}}{\gamma_1}\left(\rho_1-\gamma_1\frac{tr(\mathbf{\Sigma})}{p},\cdots,\rho_k-\gamma_k\frac{tr(\mathbf{\Sigma})}{p}\right)'$ *converges weakly to a zero-mean Gaussian vector* $\mathbf{w}=(w_1,\cdots,w_k)'$ *with* $\mathrm{var}(w_i)=\frac{2\theta}{(2i-1)^4}(1-2E(Z_{i1}^R)^2E(Z_{i1}^I)^2)$, *in which* $\theta=\lim_{p\to\infty}\frac{tr(\mathbf{\Sigma}^2)}{p}$. *When* $i\neq j$, $\mathrm{cov}(w_i,w_j)=0$.

REMARK 2. *If Assumption A7 does not hold but Assumption A6 is true, then Theorem 2.2 remains true under Assumptions A1-A3, A5 and* $\lim_{T,\ p\to\infty} \frac{p}{T} = 0$.

REMARK 3. *We now compare our results with those in [2]. [2] needs to assume that the observations are independent and that* $\mathbf{\Sigma}$ *has a spiked structure. In our paper, the observations are highly dependent. Furthermore, we need not assume a spiked structure of* $\mathbf{\Sigma}$*, since the spiked eigenvalues come naturally from the random walk structure.*

2.3. *Main results for centered sample covariance matrix* $\bar{\mathbf{B}}$. We now consider the largest eigenvalues of $\bar{\mathbf{B}}$. To characterize the limits in probability of the eigenvalues of $\bar{\mathbf{B}}$, define for $k = 1, \cdots, T$,

$$\bar{\lambda}_k = \frac{1}{2(1+\cos\bar{\theta}_k)} \quad \text{with} \quad \bar{\theta}_k = \frac{(T-k)\pi}{T}, \tag{2.14}$$

and

$$\bar{\gamma}_k = \bar{\lambda}_k \left(a_0 + 2\sum_{j=1}^{\infty} a_j (-1)^j \cos(j\bar{\theta}_k) \right). \tag{2.15}$$

We below characterize the magnitude of $\bar{\lambda}_k$ and $\bar{\gamma}_k$. The result is similar to Proposition 1.

PROPOSITION 2. *Let Assumptions A1 and A2 hold. For any fixed constant* $k \geq 1$*, there is a constant* $\bar{c}_k$ *such that*

$$\lim_{T\to\infty} \frac{\bar{\gamma}_k}{T^2} = \bar{c}_k > 0 \tag{2.16}$$

and

$$\lim_{T\to\infty} \frac{\bar{\gamma}_k}{\bar{\gamma}_1} = \lim_{T\to\infty} \frac{\bar{\lambda}_k}{\bar{\lambda}_1} = \frac{1}{k^2}. \tag{2.17}$$

We next list the results, which are similar to Theorems 2.1 and 2.2.

THEOREM 2.3. *Let Assumptions A1–A6 hold. When* $0 \leq \|\mathbf{\Pi}\|_2 = \varphi < 1$*, we obtain*

$$\|\bar{\mathbf{B}}\|_2 = O_p\left(\frac{\left(1+\sqrt{\frac{T}{p}}\right)^2}{(1-\varphi)^2} \right). \tag{2.18}$$

THEOREM 2.4. *Let Assumptions A1–A5 hold. Let* $\bar{\rho}_k$ *be the* k*th largest eigenvalue of* $\bar{\mathbf{B}}$*. Let* $\mathbf{\Pi} = \mathbf{I}$ *and* k *is fixed. We then have the following results:*

$$\frac{\bar{\rho}_k - \bar{\gamma}_k \frac{tr(\mathbf{\Sigma})}{p}}{\bar{\gamma}_1} \xrightarrow{i.p.} 0, \tag{2.19}$$

and the random vector

$$\frac{\sqrt{p}}{\bar{\gamma}_1}\left(\bar{\rho}_1 - \bar{\gamma}_1 \frac{tr(\mathbf{\Sigma})}{p}, \cdots, \bar{\rho}_k - \bar{\gamma}_k \frac{tr(\mathbf{\Sigma})}{p}\right)' \tag{2.20}$$

converges weakly to a zero-mean Gaussian vector $\bar{\mathbf{w}} = (\bar{w}_1, \cdots, \bar{w}_k)'$ *with covariance function* $\operatorname{cov}(\bar{w}_i, \bar{w}_j) = 0$ *for any* $i \neq j$ *and* $\operatorname{var}(\bar{w}_i) = \frac{2\theta}{i^4}$ *with* $\theta = \lim_{p\to\infty} \frac{tr(\mathbf{\Sigma}^2)}{p}$.

REMARK 4. *It is noted that Theorem 2.4 doesn't need Assumptions A6 and A7 due to the structure of* $\bar{\mathbf{B}}$*.*

We are now ready to introduce two new unit root tests for the high dimensional time series case before the proofs of the theorems are given in Appendix B below.

3. Unit Root Testing. This section is to explore an application of the main results to the proposal of a new unit root test for a high dimensional time series setting.

Unit root testing is to check whether time series data are nonstationary or not. Existing studies on this topic can be found in [12], [6] and [30]. In the past two decades, unit root testing in panel data has received much attention. Many researchers (see, for example, [9] and [21] for proposing the p–value based test independently, [19] for establishing the pooled t–test, and [16] for considering an averaged t–test) consider the time series case where the error process is independent across individuals. There are also some tests (see, for example, [7], [27] and [29]) proposed for the case where the error process is cross–sectional dependent. [11] also discussed subsampling hypothesis tests for nonstationary panels. [23] discussed incidental trends and the power of panel unit root tests. In Chapter 7 of a recent book, Choi [10] provided a comprehensive survey and discussion about various unit–root tests proposed for the panel data case. Meanwhile, another recent book by [28] summarized some recent developments about unit root testing for both time series and panel data settings. In the above literature, researchers often need to first estimate the covariance matrix of a panel of times associated with cross–sectional dependence. However, when the dimensionality of the time series becomes large, it is hard to consistently estimate it without imposing some structure on the covariance matrix. We therefore propose two new tests using the covariance matrices of high dimensional time series under consideration.

To this end, a key observation is that Theorems 2.2 and 2.4 indicate that the largest eigenvalues of $\mathbf{B}$ and $\bar{\mathbf{B}}$ are of order T^2 in probability (the order of γ_1 and $\bar{\gamma}_1$, which are given in Propositions 1 and 2), while Theorems 2.1, 2.3 and Assumption (A4) imply that when $0 \le \varphi < 1$, we have $\|\mathbf{B}\|_2 = o_p(T)$ and $\|\bar{\mathbf{B}}\|_2 = o_p(T)$. This motivates us to propose two new unit root tests based on the largest eigenvalues.

3.1. *The model and test statistics.* We consider the following model:

$$\mathbf{x_t} = (\mathbf{I} - \mathbf{\Pi})\phi + \mathbf{\Pi}\mathbf{x_{t-1}} + \mathbf{\Sigma}^{1/2}\mathbf{y_t}, \ 1 \le t \le T, \tag{3.1}$$

where ϕ is a p–dimensional vector. The null hypothesis H_0 is $\mathbf{\Pi} = \mathbf{I}$ and the alternative hypothesis H_1 is $\|\mathbf{\Pi}\|_2 < 1$.

Theorem 2.2 states that under $H_0: \ \mathbf{\Pi} = \mathbf{I}$, the statistic $L_p = \frac{\sqrt{p}\left(\rho_1 - \gamma_1 \frac{tr(\mathbf{\Sigma})}{p}\right)}{\gamma_1\sqrt{2\theta}}$ converges weakly to a standard normal variable. Note that $\gamma_1 \frac{tr(\Sigma)}{p}$ and $\gamma_1\sqrt{2\theta}$ are both unknown in practice. We would like to emphasize that γ_1, $\frac{tr(\mathbf{\Sigma})}{p}$ and θ can not be estimated individually. However, it is possible to estimate their product as a whole. Specifically speaking, an estimator of $\frac{\gamma_1}{\lambda_1}\frac{tr(\mathbf{\Sigma})}{p}$ is proposed as follows.

Define $\breve{\mathbf{x}}_{f,g} = (\mathbf{x_f} - \mathbf{x_{f-1}})'(\mathbf{x_g} - \mathbf{x_{g-1}})$ for $1 \le f, g \le T$. A direct calculation yields $E\breve{\mathbf{x}}_{f,g} = a_{|f-g|}tr(\mathbf{\Sigma})$. Moreover, note that $\sum_{j=1}^{m_1} a_j(-1)^j\cos(j\theta_1)$ can be approximated by $\sum_{j=1}^{m_1} a_j$ for an appropriate m_1 to be specified below. In view of this, we propose an estimator of $\frac{\gamma_1}{\lambda_1}\frac{tr(\mathbf{\Sigma})}{p}$ as

$$\mu_{m_1} = \sum_{i=2}^{T} \frac{\breve{\mathbf{x}}_{i,i}}{p(T-1)} + 2\sum_{j=1}^{m_1}\sum_{i=2}^{T-j} \frac{\breve{\mathbf{x}}_{i,i+j}}{p(T-j-1)}. \tag{3.2}$$

We next find an estimator for $\gamma_1\sqrt{2\frac{tr(\mathbf{\Sigma^2})}{p}}$. The strategy is to find an estimator for the ratio of $\gamma_1\sqrt{2\frac{tr(\mathbf{\Sigma^2})}{p}}$ and $\gamma_1\frac{tr(\mathbf{\Sigma})}{p}$ first and then construct its estimator in conjunction with μ_{m_1}, the estimator of $\frac{\gamma_1}{\lambda_1}\frac{tr(\mathbf{\Sigma})}{p}$. To this end, we first find an estimator for $a_0^2\, tr(\Sigma^2)$. One may verify that $Var(\breve{\mathbf{x}}_{f,g}) =$

$(a_{|f-g|}^2+a_0^2)tr(\mathbf{\Sigma^2})$. It is also noted that $a_{|f-g|} = o(|f-g|)$ due to Assumption A1 so that the term $a_{|f-g|}$ in $Var(\check{\mathbf{x}}_{f,g})$ can be negligible when choosing $|f-g|$ sufficiently large. We then propose an estimator for $a_0^2\, tr(\Sigma^2)$ as follows:

$$S_{\sigma^2,0} = \frac{\sum_{f=2}^{[T/2]} \sum_{g=f+[T/2]}^{T} \check{\mathbf{x}}_{f,g}^2}{(T-\frac{3}{2}[T/2])([T/2]-1)}. \tag{3.3}$$

Furthermore, one may verify that

$$\frac{\sqrt{\frac{S_{\sigma^2,0}}{p}}}{\sum_{i=2}^{T} \frac{\check{\mathbf{x}}_{i,i}}{p(T-1)}} - \frac{\sqrt{\frac{tr(\mathbf{\Sigma^2})}{p}}}{\frac{tr(\Sigma)}{p}} \xrightarrow{i.p.} 0.$$

We may then construct S_{σ^2,m_2}, the estimator of $\frac{\gamma_1}{\lambda_1}\sqrt{2\frac{tr(\mathbf{\Sigma^2})}{p}}$, as follows:

$$S_{\sigma^2,m_2} = \frac{|\mu_{m_2}|\sqrt{2\frac{S_{\sigma^2,0}}{p}}}{\sum_{i=2}^{T} \frac{\check{\mathbf{x}}_{i,i}}{p(T-1)}}$$

where m_2 is specified below.

Also, note that $\gamma_1/\lambda_1 = \bar{\gamma}_1/\bar{\lambda}_1$. Once the two estimators are available, we can construct the following test statistics, T_N and $\bar{T}_N$, of the form:

$$T_N = \sqrt{p}\frac{\rho_1 - \lambda_1\mu_{m_1}}{\lambda_1 S_{\sigma^2,m_2}} \tag{3.4}$$

and

$$\bar{T}_N = \sqrt{p}\frac{\bar{\rho}_1 - \bar{\lambda}_1\mu_{m_1}}{\bar{\lambda}_1 S_{\sigma^2,m_2}}, \tag{3.5}$$

where λ_1 and $\bar{\lambda}_1$ are given in (2.5) and (2.14), respectively. Let $[x]$ stand for the largest integer part of x.

THEOREM 3.1. *Let Assumptions A1–A5 hold,* $m_1 = [\sqrt{p}]$ *and* m_2 *tends to infinity. Under* $H_0:\ \mathbf{\Pi} = \mathbf{I}$*, we have*

$$\bar{T}_N \xrightarrow{d} N(0,1), \tag{3.6}$$

where $\xrightarrow{d}$ *stands for convergence in distribution.*

Furthermore, if Assumptions A7 also holds, under $H_0:\ \mathbf{\Pi} = \mathbf{I}$*, we have*

$$T_N \xrightarrow{d} N(0,1). \tag{3.7}$$

Remark 5. *The conditions imposed on m_1 and m_2 can be further relaxed. For example, if there exists a positive integer s such that $b_i = 0$ for any $i > s$ in (2.1), we find $a_i = 0$ for any $i > s$ in (2.7). So one can choose $m_1 = m_2 = \min\{s, [\sqrt{p}]\}$ in this case. This point helps us to simplify the design and the verifications of the assumptions for the simulation in Section 4 below.*

Now we investigate the power of T_N and $\bar{T}_N$ for the case where $\{Y_{tj}\}$ in (2.1) are i.i.d..

Theorem 3.2. *Let Assumptions A1–A5 hold with $b_i = 0$ for $i \geq 1$. Consider $H_1 :\ \mathbf{\Pi} = \varphi \mathbf{I}$ for $0 \leq \varphi < 1$. Then under the case of $m_1 = m_2 = 0$, we have*

$$\lim_{T\to\infty} P\left(\bar{T}_N > C_0 | H_1\right) = 1 \tag{3.8}$$

for some $C_0 > \ell_\alpha$, where ℓ_α is the α–level critical value of the standard normal distribution.

Furthermore, if $\|\phi\|_2^2 = O(p)$, then

$$\lim_{T\to\infty} P\left(T_N > C_0 | H_1\right) = 1. \tag{3.9}$$

Remark 6. *Although $\bar{T}_N$ and T_N may have the same asymptotic results when p and T are big enough, there may be differences under the small sample case. In fact under H_0, $\mathbf{x_0}$ affects the largest eigenvalues of $\mathbf{B}$ but doesn't affect the largest eigenvalues of $\bar{\mathbf{B}}$. So it may affect the size of T_N when the sample is small. Under H_1, ϕ affects the largest eigenvalues of $\mathbf{B}$ but doesn't affect the largest eigenvalues of $\bar{\mathbf{B}}$. It may affect the power of T_N when the sample is small. So $\bar{T}_N$ may be more useful than T_N when we don't have ϕ or $\mathbf{x_0}$. But when we have the condition that $\phi = 0$ and $\mathbf{x_0} = 0$, $\gamma_1 \approx 4\bar{\gamma}_1$ so that T_N can have a stronger power than $\bar{T}_N$ under small sample cases.*

Remark 7. *There are some well known panel unit root tests (e.g. [9] and [19]). They considered the case of $\mathbf{\Pi} = \mathrm{diag}(\varphi_1, \cdots, \varphi_N)$ and used the estimators of φ_i to test whether $\mathbf{\Pi} = \mathbf{I}$. Moreover, when the covariance matrix $\mathbf{\Sigma}$ is involved, it has to be estimated in order to test whether $\mathbf{\Pi} = \mathbf{I}$ (e.g. [7]). So such existing tests may only work for the finite-dimensional case. By contrast, our test makes the best use of the properties of the largest eigenvalues of B instead of estimating φ_i. In addition, we do not impose special structures, such as sparsity on the covariance matrix $\mathbf{\Sigma}$.*

Before the proofs of Theorems 3.1–3.2 are given in Appendix B, we evaluate the finite sample performance of the proposed tests and also compare them with two natural competitors in Section 4 below.

4. Simulation. This section is to conduct some simulations to investigate the size and power of T_N and $\bar{T}_N$.

4.1. *The selection of* m_1 *and* m_2*.* Recalling Remark 5, we below propose a method to choose suitable m_1 and m_2. Note that

$$\zeta_j = \frac{\sum_{i=2}^{T-j} \frac{\check{\mathbf{x}}_{i,i+j}}{p(T-j-1)}}{\sum_{i=2}^{T} \frac{\check{\mathbf{x}}_{i,i}}{p(T-1)}} \xrightarrow{i.p.} \frac{a_j}{a_0}$$

with the rate $\frac{1}{\sqrt{pT}}$. Particularly $\zeta_j = O(\frac{1}{\sqrt{pT}})$ if $a_j = 0$. Moreover, if there exists a positive integer s such that $b_i = 0$ for any $i > s$ in (2.1), we find $a_i = 0$ for any $i > s$ in (2.7). So one can choose $m_1 = m_2 = \min\{s, [\sqrt{p}]\}$ in this case. In practice one can see whether $a_j = 0$ by comparing ζ_j with $p^{-1/2}T^{-1/4}$. Here $p^{-1/2}T^{-1/4}$ is used as a bound instead of $\frac{1}{\sqrt{pT}}$ since the convergence rate of μ_{m_1} to $\frac{\gamma_1}{\lambda_1}\frac{tr\Sigma}{p}$ should be $o(p^{-1/2})$. In view of this, we propose the following way of selecting m_1 and m_2:

(4.1)
$$\hat{m}_1 = \hat{m}_2 = \min\{\{0 \le i < [\sqrt{p}] : |\zeta_j| < p^{-1/2}T^{-1/4}, i < j < [\sqrt{p}]\} \cup \{[\sqrt{p}]\}\}.$$

Note that $\hat{m}_1$ and $\hat{m}_2$ work well when p and T are big enough. While when p and T are small, $\hat{m}_1$ and $\hat{m}_2$ may be affected by $\frac{a_j}{a_0}$. If $a_j \neq 0$ but $\frac{a_j}{a_0}$ is small, $\hat{m}_1$ and $\hat{m}_2$ may cause some problem when p and T are small.

4.2. *The parametric bootstrap method.* We also consider a parametric bootstrap method for our test statistics T_N and $\bar{T}_N$. Let $\dot{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{x_t} - \mathbf{x_{t-1}})(\mathbf{x_t} - \mathbf{x_{t-1}})'$. If there is a constant $\dot{C} > 0$ such that $\frac{p}{T} \le \dot{C}$, we can find that $\| \dot{\mathbf{\Sigma}} \|_2 = O_p(1)$ and $\frac{tr(\dot{\mathbf{\Sigma}})}{p} = \dot{M}_1 + O_p(\frac{1}{\sqrt{p}})$, where $\dot{M}_1 > 0$. It is easily seen that Assumption A3 still holds for $\dot{\mathbf{\Sigma}}$. We then draw a new sample $\dot{\mathbf{x}}_\mathbf{t} = \dot{\mathbf{x}}_\mathbf{t-1} + \dot{\mathbf{\Sigma}}^{1/2}\dot{\mathbf{y}}_\mathbf{t}$ where $\dot{\mathbf{y}}_\mathbf{t}$ is a p-dimensional random vector from $N(0, \mathbf{I_p})$ and $\dot{\mathbf{y}}_\mathbf{t}$ is independent over t. Note that Assumptions A1–A7 still hold for $\dot{\mathbf{x}}_\mathbf{t}$. Let $\dot{\mathbf{X}} = (\dot{\mathbf{x}}_\mathbf{1}, \cdots, \dot{\mathbf{x}}_\mathbf{T})'$. We define $\dot{T}_N$ and $\dot{\bar{T}}_N$ from $\dot{\mathbf{X}}$, the analogues of T_N and $\bar{T}_N$, respectively. It follows from Theorem 3.1 that $\dot{T}_N \xrightarrow{d} N(0,1)$ and $\dot{\bar{T}}_N \xrightarrow{d} N(0,1)$. So for any p and T we can redraw $\dot{\mathbf{x}}_\mathbf{t}$ for many times (e.g. 200 times) to get an empirical distributions for each of

$\dot{T}_N$ and $\dot{\bar{T}}_N$. Then we use the critical values from the empirical distributions to replace the critical values calculated from $N(0,1)$. When p and T are not big, the simulations show that $\dot{T}_N$ and $\dot{\bar{T}}_N$ based on the critical values from the empirical distributions perform better than these tests associated with the corresponding critical values calculated from $N(0,1)$.

4.3. *Comparison with the existing tests.* There are several existing unit root tests available for panel data. Some of them consider the case where there is no cross-sectional dependence (see, for example, the IPS test proposed in [16]). If there is cross-sectional dependence, the IPS test doesn't work. To test for nonstationarity in the panel data case with cross-sectional dependence, [7] showed that the Bootstrap method with estimation of $\mathbf{\Sigma}$ performs better for the case where p is fixed and T is large. [7] also stated that the Bootstrap-OLS performs better than Bootstrap-GLS when p is large. Furthermore, GLS doesn't work when $p \geq T$. We therefore compare T_N with the t–statistic corresponding to the Bootstrap-OLS t^*_{ols} and the F-statistic corresponding to Bootstrap-OLS F^*_{ols}.

We use the setting $\mathbf{y_t} = \mathbf{z_t}$ and $\mathbf{\Sigma} = \left(\Sigma_{i,j}\right) = \left(0.3^{|i-j|}\right)$. We compare the size performance of our test T_N with the two tests t^*_{ols} and F^*_{ols} under H_0 with $\mathbf{x_0} = 0$ and $\phi = 0$. Table 1 reports the results of the three tests based on 1000 replications, 500 bootstrap replications and different values of p and T. The nominal size throughout this section is set to be 0.05.

Table 1
The empirical size of three tests

the test	T \ p	5	10	20	40	60	80
T_N	40	0.057	0.057	0.041	0.048	0.050	0.040
t^*_{ols}	40	0.045	0.028	0.014	0.000	0.000	0.000
F^*_{ols}	40	0.054	0.044	0.027	0.000	0.003	0.001
T_N	60	0.053	0.050	0.048	0.055	0.048	0.044
t^*_{ols}	60	0.046	0.031	0.016	0.003	0.000	0.000
F^*_{ols}	60	0.044	0.047	0.024	0.007	0.000	0.002
T_N	80	0.053	0.048	0.041	0.052	0.048	0.041
t^*_{ols}	80	0.045	0.033	0.023	0.006	0.000	0.000
F^*_{ols}	80	0.064	0.035	0.027	0.011	0.003	0.000

Then, we compare our test $\bar{T}_N$ with the two tests t^*_{ols} and F^*_{ols} under H_0 and $\mathbf{x_0} = 0, \mathbf{\Sigma} = \left(\Sigma_{i,j}\right) = \left(\frac{1}{(i-j)^2+1}\right)$. We sample each element of ϕ from the standard normal distribution. The results of the three test statistics based on 1000 replications, 500 bootstrap replications and different values of p and T are reported in Table 2.

TABLE 2
The empirical size of three tests

the test	T \ p	5	10	20	40	60	80
$\bar{T}_N$	40	0.070	0.056	0.062	0.052	0.038	0.043
t^*_{ols}	40	0.036	0.013	0.008	0.001	0.000	0.000
F^*_{ols}	40	0.056	0.029	0.014	0.001	0.000	0.000
$\bar{T}_N$	60	0.061	0.060	0.047	0.041	0.045	0.053
t^*_{ols}	60	0.041	0.037	0.011	0.001	0.000	0.000
F^*_{ols}	60	0.052	0.054	0.027	0.002	0.000	0.000
$\bar{T}_N$	80	0.055	0.058	0.053	0.048	0.041	0.048
t^*_{ols}	80	0.041	0.043	0.015	0.006	0.000	0.000
F^*_{ols}	80	0.041	0.045	0.035	0.008	0.000	0.000

One can observe that when p becomes large, both t^*_{ols} and F^*_{ols} have a poor size property even though $\mathbf{y_t}$ is independent over t. This indicates that their asymptotic distributions may not hold under the null hypothesis when p is large. One of the reasons is that when p is large and the population covariance matrix does not have any special structures, we cannot find any consistent estimates for the population covariance matrix and the unknown parameters involved. As a consequence, their asymptotic distributions may fail to hold under the null.

4.4. *Simulation results for T_N under an MA(1) model.* We now consider the setting where $\mathbf{y_t} = \psi\mathbf{z_{t-1}} + \mathbf{z_t}$, $\psi = 0.5$ and $\mathbf{\Sigma} = \Big(\Sigma_{i,j}\Big) = \Big(0.3^{|i-j|}\Big)$. To show the performance with the non–diagonal $\mathbf{\Pi}$, we design the following matrix as an alternative one:

$$(\mathbf{\Pi_2})_{ij} = \begin{cases} 0.5 & i = j, \\ 0.2 & |i-j| = 1, \\ 0 & |i-j| \geq 2. \end{cases}$$

We consider the performance of T_N and set $\phi = 0$. Under H_0 we set $\mathbf{x_0} = 0$. Under H_1 we generate the data by (3.1) with $t = -51, -50, \cdots, T$. Using an asymptotic critical value calculated from $N(0,1)$, the size and power results of T_N based on 1000 replications and different values of p, T and $\mathbf{\Pi}$ are reported in Table 3. We also use the parametric bootstrap method proposed in Section 4.2. The size and power results of T_N based on 1000 replications, 200 bootstrap replications and different values of p, T and $\mathbf{\Pi}$ are reported in Table 4.

4.5. *Simulation results for $\bar{T}_N$ under an MA(1) model.* We still use the setting in Section 4.4 but sample each element of ϕ from the standard normal

Table 3
The results for T_N and MA(1)

p	T	**I** (size)	**0.95I** (power)	**0.9I** (power)	Π_2 (power)
20	20	0.019	0.102	0.216	0.510
20	30	0.037	0.109	0.672	0.830
20	40	0.036	0.346	0.951	0.935
20	60	0.043	0.896	1.000	0.997
20	80	0.039	0.997	1.000	1.000
40	20	0.019	0.102	0.580	0.710
40	30	0.028	0.301	0.964	0.938
40	40	0.031	0.752	0.999	0.974
40	60	0.034	0.997	1.000	0.998
40	80	0.033	1.000	1.000	1.000
60	20	0.021	0.100	0.766	0.876
60	30	0.029	0.421	0.998	0.981
60	40	0.033	0.905	1.000	0.989
60	60	0.045	1.000	1.000	0.998
60	80	0.046	1.000	1.000	1.000
80	20	0.020	0.116	0.870	0.932
80	30	0.029	0.561	1.000	0.996
80	40	0.032	0.966	1.000	0.997
80	60	0.036	1.000	1.000	1.000
80	80	0.034	1.000	1.000	1.000

Table 4
The results for T_N and MA(1) with the parametric bootstrap method

p	T	**I** (size)	**0.95I** (power)	**0.9I** (power)	Π_2 (power)
20	20	0.031	0.144	0.636	0.812
20	30	0.063	0.464	0.974	0.936
20	40	0.051	0.818	0.998	0.992
20	60	0.049	0.992	1.000	1.000
20	80	0.082	1.000	1.000	1.000
40	20	0.061	0.140	0.860	0.838
40	30	0.051	0.578	0.990	0.972
40	40	0.041	0.926	1.000	0.990
40	60	0.052	0.998	1.000	1.000
40	80	0.054	1.000	1.000	1.000
60	20	0.055	0.126	0.930	0.932
60	30	0.048	0.676	1.000	0.994
60	40	0.068	0.972	1.000	0.996
60	60	0.053	1.000	1.000	1.000
60	80	0.056	1.000	1.000	1.000
80	20	0.055	0.132	0.950	0.960
80	30	0.047	0.742	1.000	0.994
80	40	0.056	0.984	1.000	0.996
80	60	0.054	1.000	1.000	1.000
80	80	0.057	1.000	1.000	1.000

distribution. In each case, we use the critical value calculated from either $N(0,1)$ or by the parametric bootstrap method. The size and power results of $\bar{T}_N$ based on 1000 replications and different values of p, T and $\mathbf{\Pi}$ are reported in Table 6.

TABLE 5
The results for $\bar{T}_N$ and MA(1)

p	T	**I** (size)	**0.95I** (power)	**0.9I** (power)	Π_2 (power)
20	20	0.018	0.018	0.013	0.100
20	30	0.042	0.029	0.124	0.213
20	40	0.043	0.071	0.290	0.383
20	60	0.046	0.264	0.746	0.733
20	80	0.055	0.580	0.959	0.907
40	20	0.016	0.034	0.075	0.176
40	30	0.033	0.081	0.290	0.363
40	40	0.034	0.235	0.584	0.572
40	60	0.044	0.708	0.985	0.919
40	80	0.043	0.968	1.000	0.987
60	20	0.014	0.036	0.144	0.254
60	30	0.029	0.202	0.523	0.518
60	40	0.036	0.408	0.823	0.729
60	60	0.039	0.870	0.999	0.936
60	80	0.042	0.993	1.000	0.998
80	20	0.012	0.064	0.191	0.310
80	30	0.032	0.267	0.661	0.644
80	40	0.037	0.532	0.934	0.800
80	60	0.043	0.945	1.000	0.971
80	80	0.039	0.997	1.000	1.000

When p is small, the size and power results of T_N and $\bar{T}_N$ based on the critical value either calculated from $N(0,1)$ or by the bootstrap method are reported in Tables 7 and 8. From Tables 7 and 8, one can observe that while $\bar{T}_N$ and T_N roughly have similar size values, the power of T_N is slightly better than that of $\bar{T}_N$. The power of the statistics of $\bar{T}_N$ and T_N improves when p and T increase. The parametric bootstrap (proposed in this paper) based critical value in each case results in a stable size and better power than using an asymptotic critical value for the case where p is as small as $p = 5$ or $p = 10$.

In summary, for the case of $p = 5$ or $p = 10$, Tables 7 and 8 show that the size and power values of T_N and $\bar{T}_N$ based on the asymptotic critical value of $N(0,1)$ are much less stable and reasonable than those based on the parametric bootstrap critical value in each case. Tables 3–6 then show that when $p \geq 20$ and $T \geq 20$, there are stable sizes and reasonable power values for both T_N and $\bar{T}_N$ based on 1000 replications, 200 bootstrap replications

Table 6
The results for $\bar{T}_N$ and MA(1) with the parametric bootstrap method

p	T	**I** (size)	**0.95I** (power)	**0.9I** (power)	Π_2 (power)
20	20	0.034	0.144	0.239	0.326
20	30	0.051	0.310	0.606	0.596
20	40	0.061	0.502	0.837	0.782
20	60	0.067	0.824	0.986	0.971
20	80	0.078	0.946	1.000	0.996
40	20	0.040	0.188	0.352	0.412
40	30	0.049	0.412	0.695	0.660
40	40	0.045	0.604	0.873	0.782
40	60	0.054	0.950	0.999	0.972
40	80	0.053	0.994	1.000	0.998
60	20	0.034	0.232	0.452	0.504
60	30	0.049	0.506	0.807	0.704
60	40	0.047	0.728	0.961	0.850
60	60	0.048	0.980	1.000	0.980
60	80	0.062	1.000	1.000	0.998
80	20	0.033	0.276	0.512	0.534
80	30	0.040	0.548	0.903	0.826
80	40	0.046	0.816	0.986	0.910
80	60	0.052	0.990	1.000	0.986
80	80	0.064	1.000	1.000	0.998

Table 7
The results for T_N and small p

p	T	critical value	**I** (size)	**0.95I** (power)	**0.9I** (power)	Π_2 (power)
5	20	N(0,1)	0.040	0.056	0.008	0.070
		bootstrap	0.084	0.114	0.308	0.560
5	30	N(0,1)	0.039	0.014	0.010	0.038
		bootstrap	0.077	0.180	0.580	0.762
5	40	N(0,1)	0.051	0.002	0.012	0.030
		bootstrap	0.079	0.262	0.772	0.882
5	60	N(0,1)	0.055	0.000	0.006	0.002
		bootstrap	0.079	0.570	0.938	0.986
5	80	N(0,1)	0.049	0.000	0.002	0.002
		bootstrap	0.076	0.816	0.992	0.992
10	20	N(0,1)	0.023	0.078	0.048	0.202
		bootstrap	0.085	0.132	0.462	0.664
10	30	N(0,1)	0.031	0.016	0.142	0.330
		bootstrap	0.075	0.240	0.826	0.896
10	40	N(0,1)	0.039	0.042	0.322	0.416
		bootstrap	0.077	0.580	0.972	0.966
10	60	N(0,1)	0.037	0.126	0.558	0.502
		bootstrap	0.069	0.894	1.000	0.998
10	80	N(0,1)	0.049	0.246	0.678	0.598
		bootstrap	0.064	0.982	1.000	1.000

TABLE 8
The results for $\bar{T}_N$ and small p

p	T	critical value	$\mathbf{I}$(size)	$\mathbf{0.95I}$(power)	$\mathbf{0.9I}$(power)	Π_2(power)
5	20	N(0,1)	0.031	0.014	0.008	0.024
		bootstrap	0.046	0.086	0.155	0.238
5	30	N(0,1)	0.039	0.002	0.002	0.006
		bootstrap	0.066	0.132	0.284	0.422
5	40	N(0,1)	0.051	0.000	0.000	0.002
		bootstrap	0.071	0.170	0.417	0.518
5	60	N(0,1)	0.050	0.000	0.000	0.002
		bootstrap	0.063	0.328	0.712	0.754
5	80	N(0,1)	0.053	0.000	0.000	0.000
		bootstrap	0.069	0.466	0.896	0.926
10	20	N(0,1)	0.025	0.004	0.009	0.081
		bootstrap	0.049	0.098	0.218	0.308
10	30	N(0,1)	0.043	0.002	0.018	0.068
		bootstrap	0.056	0.214	0.450	0.471
10	40	N(0,1)	0.046	0.014	0.031	0.108
		bootstrap	0.073	0.300	0.653	0.684
10	60	N(0,1)	0.062	0.022	0.117	0.168
		bootstrap	0.069	0.560	0.904	0.880
10	80	N(0,1)	0.057	0.022	0.217	0.234
		bootstrap	0.075	0.748	0.991	0.960

and different values of p, T and $\mathbf{\Pi}$.

REMARK 8. *In Tables 3–8, one can find that using the bootstrap critical values also leads to better empirical power. The reason is that T_N and $\bar{T}_N$ under H_1 have the order $O(\sqrt{p}(1-\frac{C(1+\sqrt{\frac{T}{p}})^2}{T^2}))$ so that the values of T_N and $\bar{T}_N$ under H_1 are not very big when p or T is small. So the change of the critical value may influence the power very much.*

5. Conclusions and Discussion. This paper has developed an asymptotic theory for the largest eigenvalues of the covariance matrix of a high dimensional time series vector. As an application, a new unit root test developed for testing nonstationarity in high dimensional time series vectors has been proposed and then discussed both theoretically and numerically. The small sample properties discussed in Section 4 have offered the support to the theory established in Sections 2 and 3.

One possible extension involves the case where either a deterministic trending time series component or a factor model structure is included in model (3.1). As a consequence, it may be more appropriate to compare the corresponding versions of T_N and $\bar{T}_N$ with those proposed by [16], [27] and

[29]. As suggested by the referees, another extension of model (3.1) is to take into account certain type of cointegrating structures. Appendix D of the supplementary document gives some brief discussion about possible extensions, which require developing new techniques and should be left for future research.

6. Acknowledgments. We would like to thank Professor Yoosoon Chang for kindly providing us with the computer code for the simulations in [7].

APPENDIX A: RESULTS FOR TRUNCATED MATRICES

This section is to consider the truncated version of the sample covariance matrix. Let $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_T)'$ be a $T \times p$ random matrix. Define

$$Y_{ij,l} = \sum_{k=0}^{l} b_k Z_{i-k,j}$$

with $l = \max\{p, T\}$, a truncated version of Y_{tj} in (2.1). However, to simplify notation, we let $b_i = 0$ for all $i > l$ in this section, so that we can still use Y_{ij} instead of $Y_{ij,l}$. In this way a_i defined in (2.7) and Y_{tj} in (2.1) respectively become

$$a_i = \sum_{k=0}^{l-i} b_k b_{k+i}, \qquad Y_{tj} = \sum_{k=0}^{l} b_k Z_{t-k,j}.$$

Furthermore let $\mathbf{F} = (F_{ij})$ be a $T \times (T + l)$ matrix with

$$F_{ij} = \begin{cases} b_{l+i-j} & i \le j \le i + l, \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

It follows that $\mathbf{Y} = \mathbf{F}\mathbf{Z_p}$, where $\mathbf{Z_p}$ is a $(T+l)\times p$ random matrix with $(\mathbf{Z_p})_{i,j} = Z_{i-l,j}$. For the sake of notation simplicity, we below denote $\mathbf{Z_p}$ by $\mathbf{Z}$ and $(\mathbf{Z_p})_{i,j}$ by Z_{ij}. Let $\mathbf{A} = (A_{ij})_{T\times T} = (a_{|i-j|})_{T\times T}$. We then have $\mathbf{A} = \mathbf{F}\mathbf{F}'$. We would like to remind the readers that l depends on T, so that $a_{|i-j|}$ depends on T.

We also assume that $\mathbf{x_0} = \mathbf{0}$ in this section.

A.1. Upper bound of the spectral norm of B for stationary data.

This subsection is to investigate the upper bound of the spectral norm of $\mathbf{B}$ for stationary data.

Proposition 3. *Suppose that Assumptions A1-A5 hold. When* $0 \le \|\mathbf{\Pi}\|_2 = \varphi < 1$,

$$\lim_{T\to\infty} P\left(\|\mathbf{B}\|_2 \le \frac{8\sum_{i\ge 0}|a_i|}{(1-\varphi)^2} M_0 \left(1 + \sqrt{\frac{T}{p}}\right)^2\right) = 1.$$

The proof of the proposition is available from the supplementary file.

A.2. Convergence in Probability and CLT of the first k largest eigenvalues when $\mathbf{\Pi} = \mathbf{I}$. Define $\mathbf{C} = (C_{ij})_{1\le i,j\le T}$ to be a $T \times T$ lower triangular matrix with

$$C_{ij} = 0 \ \text{ for } j > i \text{ and } \ C_{ij} = 1 \ \text{ for } 1 \le j \le i. \tag{A.2}$$

In this case one has

$$\mathbf{B} = (1/p)\mathbf{X}\mathbf{X}^* = (1/p)\mathbf{C}\mathbf{Y}\mathbf{\Sigma}\mathbf{Y}^*\mathbf{C}^* = (1/p)\mathbf{C}\mathbf{F}\mathbf{Z_p}\mathbf{\Sigma}\mathbf{Z_p^*}\mathbf{F}^*\mathbf{C}^*. \tag{A.3}$$

PROPOSITION 4. *Suppose that Assumptions A1-A5 hold. Let ρ_k be the kth largest eigenvalue of $\mathbf{B}$. When $\mathbf{\Pi} = \mathbf{I}$, $\frac{\rho_k - \gamma_k \frac{tr(\mathbf{\Sigma})}{p}}{\gamma_1} \to 0$ in probability.*

PROPOSITION 5. *Suppose that Assumptions A1-A5 hold. Let ρ_k be the kth largest eigenvalue of $\mathbf{B}$. When $\mathbf{\Pi} = \mathbf{I}$, $(\sqrt{p}\frac{\rho_1-\gamma_1}{\gamma_1}, \cdots, \sqrt{p}\frac{\rho_k-\gamma_k}{\gamma_1})'$ converges weakly to a zero-mean Gaussian vector $\mathbf{w} = (w_1, \cdots, w_k)'$ with covariance $cov(w_i, w_j) = \delta_{ij}\frac{\theta}{(2i-1)^4}(2 - 4E(Z_{i1}^R)^2 E(Z_{i1}^I)^2)$ and $\theta = \lim_{p\to\infty} \frac{tr(\mathbf{\Sigma}^2)}{p}$.*

The proofs of the propositions are available from the supplementary file.

A.3. The results for $\bar{\mathbf{B}}$. The following results for $\bar{\mathbf{B}}$ are similar to those for $\mathbf{B}$. In view of (A.3), write

$$\bar{\mathbf{B}} = (1/p)\mathbf{H}\mathbf{C}\mathbf{F}\mathbf{Z_p}\mathbf{\Sigma}\mathbf{Z_p^*}\mathbf{F}^*\mathbf{C}^*\mathbf{H}^*, \tag{A.4}$$

where $\mathbf{H} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{T}$ with the $p \times 1$ vector $\mathbf{1}$ consisting of all one.

PROPOSITION 6. *Suppose that Assumptions A1-A5 hold. Let $\bar{\rho}_k$ be the kth largest eigenvalue of $\bar{\mathbf{B}}$. When $\mathbf{\Pi} = \mathbf{I}$, $\frac{\bar{\rho}_k - \bar{\gamma}_k \frac{tr(\mathbf{\Sigma})}{p}}{\bar{\gamma}_1} \to 0$ in probability.*

PROPOSITION 7. *Suppose that Assumptions A1-A5 hold. Let $\bar{\rho}_k$ be the kth largest eigenvalue of $\bar{\mathbf{B}}$. When $\mathbf{\Pi} = \mathbf{I}$, $(\sqrt{p}\frac{\bar{\rho}_1-\bar{\gamma}_1}{\bar{\gamma}_1}, \cdots, \sqrt{p}\frac{\bar{\rho}_k-\bar{\gamma}_k}{\bar{\gamma}_1})'$ converges weakly to a zero-mean Gaussian vector $\bar{\mathbf{w}} = (\bar{w}_1, \cdots, \bar{w}_k)'$ with covariance $cov(\bar{w}_i, \bar{w}_j) = \delta_{ij}\frac{\theta}{i^4}(2 - 4E(Z_{i1}^R)^2 E(Z_{i1}^I)^2)$ and $\theta = \lim_{p\to\infty} \frac{tr(\mathbf{\Sigma}^2)}{p}$.*

The proofs of the propositions are available from the supplementary file.

APPENDIX B: PROOFS OF THE MAIN RESULTS

This section is to prove that the results obtained in Section 4 still hold for the general linear process (without the truncation step performed there) and the general initial vector $\mathbf{x}_0$. We define a $T \times p$ matrix $\mathbf{X_0} = (\mathbf{x_0}, \cdots, \mathbf{x_0})'$ consisting of the initial vector $\mathbf{x}_0$ of the time series. When $\mathbf{\Pi} = \mathbf{I}$, we may rewrite $\mathbf{X} = \mathbf{C}\mathbf{Y}\mathbf{\Sigma}^{1/2} + \mathbf{X_0}$

and $\bar{\mathbf{X}} = \frac{\mathbf{1}\mathbf{1}'}{T}\mathbf{C}\mathbf{Y}\mathbf{\Sigma}^{1/2} + \mathbf{X_0}$ so that the sample covariance matrices $\mathbf{B}$ and $\bar{\mathbf{B}}$ can be rewritten as follows:

$$\mathbf{B} = \frac{1}{p}\mathbf{X}\mathbf{X}^* = \frac{1}{p}\mathbf{C}\mathbf{Y}\mathbf{\Sigma}\mathbf{Y}^*\mathbf{C}^* + \frac{1}{p}\mathbf{C}\mathbf{Y}\mathbf{\Sigma}^{1/2}\mathbf{X_0}^* + \frac{1}{p}\mathbf{X_0}\mathbf{\Sigma}^{1/2}\mathbf{Y}^*\mathbf{C}^* + \frac{1}{p}\mathbf{X_0}\mathbf{X_0}^* \tag{B.1}$$

and

$$\bar{\mathbf{B}} = \frac{1}{p}(\mathbf{X}-\bar{\mathbf{X}})(\mathbf{X}-\bar{\mathbf{X}})^* = \frac{1}{p}(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{T})\mathbf{C}\mathbf{Y}\mathbf{\Sigma}\mathbf{Y}^*\mathbf{C}^*(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{T})^*. \tag{B.2}$$

Lemma 1. *Recall the definitions of* $\mathbf{Y}$, λ_k *and* γ_k *in Section 2. Let* $l = \max\{p, T\}$ *and* $\mathbf{Y_l}$ *be the truncated matrix of* $\mathbf{Y}$ *in Section 4. Define*

$$\gamma_{k,l} = \lambda_k(a_{0,l} + 2\sum_{1\le j\le T-1} a_{j,l}(-1)^j\cos(j\theta_k))$$

where

$$a_{j,l} = \sum_{j\le k\le l} b_k b_{k-j}. \tag{B.3}$$

Then when $\mathbf{\Pi} = \mathbf{I}$,

$$\left|\left|\frac{(1/p)\mathbf{C}(\mathbf{Y}\mathbf{\Sigma}\mathbf{Y}^* - \mathbf{Y_l}\mathbf{\Sigma}\mathbf{Y_l}^*)\mathbf{C}^*}{\gamma_{1,l}}\right|\right|_2 = o_p(p^{-1/2}) \tag{B.4}$$

and

$$\frac{|\gamma_{k,l} - \gamma_k|}{\gamma_{1,l}} = o(1). \tag{B.5}$$

Proof of Lemma 1. We consider (B.5) first. To this end, observe that Assumption (A1) implies that

$$\sum_{i=0}^{\infty} i|a_i| < \infty, \tag{B.6}$$

because

$$\sum_{i=0}^{\infty} i|a_i| \le \sum_{i=0}^{\infty} i\sum_{k=0}^{\infty}|b_k||b_{k+i}| = \sum_{k=0}^{\infty}|b_k|(\sum_{i=0}^{\infty} i|b_{k+i}|) \le \sum_{k=0}^{\infty}|b_k|(\sum_{i=0}^{\infty} i|b_i|).$$

Write

$$\frac{|\gamma_{k,l} - \gamma_k|}{\gamma_{1,l}} \le \frac{\lambda_k}{\gamma_{1,l}}\left(\sum_{k>l} b_k^2 + 2\sum_{j=1}^{T-1}\sum_{k>l}|b_k||b_{k-j}| + 2\sum_{j\ge T}|a_j|\right)$$

$$\le \frac{\lambda_k}{\gamma_{1,l}}\left(\sum_{k>l} b_k^2 + 2\sum_{j=1}^{\infty}|b_j|\sum_{k>l}|b_k| + 2\sum_{j\ge T}|a_j|\right).$$

From (B.6) and Assumption (A1), we obtain that

$$\sum_{k>l} b_k^2 + 2\sum_{j=1}^{\infty} |b_j| \sum_{k>l} |b_k| + 2\sum_{j\geq T} |a_j| = o(1).$$

Moreover, Lemma C.2 and Assumption (A1) (or (C.12)) imply that $\frac{\lambda_k}{\gamma_{1,l}}$ is bounded. So we conclude (B.5).

Now, we consider (B.4). Using Lemma C.1 in the supplementary file, observe that

$$\begin{aligned}\left\|\frac{(1/p)\mathbf{C}\left(\mathbf{Y}\boldsymbol{\Sigma}\mathbf{Y}^* - \mathbf{Y_1}\boldsymbol{\Sigma}\mathbf{Y_1}^*\right)\mathbf{C}^*}{\gamma_{1,l}}\right\|_2 &\leq \frac{\parallel \mathbf{C} \parallel_2^2}{\gamma_{1,l}} \parallel (1/p)(\mathbf{Y}\boldsymbol{\Sigma}\mathbf{Y}^* - \mathbf{Y_1}\boldsymbol{\Sigma}\mathbf{Y_1}^*) \parallel_2 \\ &= \frac{\lambda_1}{\gamma_{1,l}} \parallel (1/p)(\mathbf{Y}\boldsymbol{\Sigma}\mathbf{Y}^* - \mathbf{Y_1}\boldsymbol{\Sigma}\mathbf{Y_1}^*) \parallel_2 .\end{aligned}$$

As before $\frac{\lambda_1}{\gamma_{1,l}}$ is bounded. So we just need to consider $\parallel (1/p)(\mathbf{Y}\boldsymbol{\Sigma}\mathbf{Y}^* - \mathbf{Y_1}\boldsymbol{\Sigma}\mathbf{Y_1}^*) \parallel_2$. Let $\mathbf{K} = (K_{ij})_{1\leq i\leq T, 1\leq j\leq p} = \mathbf{Y} - \mathbf{Y_1}$. We can obtain that $K_{ij} = \sum_{k=l+1}^{\infty} b_k Z_{i-k,j}$ and

$$E|K_{ij}|^2 = \sum_{k=l+1}^{\infty} b_k^2.$$

By Assumption (A1), we can get

$$E|K_{ij}|^2 = \sum_{k=l+1}^{\infty} b_k^2 \leq l^{-2} \sum_{k=l+1}^{\infty} k^2 |b_k|^2 = o(l^{-2}),$$

which implies

$$E\left\|\frac{1}{\sqrt{p}}\mathbf{K}\right\|_{F^2} = o(Tl^{-2}).$$

This, together with (C.2), implies that

(B.7)
$$\begin{aligned}&\parallel (1/p)(\mathbf{Y}\boldsymbol{\Sigma}\mathbf{Y}^* - \mathbf{Y_1}\boldsymbol{\Sigma}\mathbf{Y_1}^*) \parallel_2 = \parallel (1/p)(\mathbf{K}\boldsymbol{\Sigma}\mathbf{Y_1}^* + \mathbf{Y_1}\boldsymbol{\Sigma}\mathbf{K}^* + \mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^*) \parallel_2 \\ &\leq 2 \parallel \frac{1}{\sqrt{p}}\mathbf{K} \parallel_F \|\boldsymbol{\Sigma}\|_2 \parallel \frac{1}{\sqrt{p}}\mathbf{Y_1} \parallel_2 + \parallel \frac{1}{\sqrt{p}}\mathbf{K} \parallel_F^2 \|\boldsymbol{\Sigma}\|_2 = o_p(p^{-1/2}).\end{aligned}$$

This concludes (B.4). □

Proof of Theorem 2.2. At first we prove (2.11). Recalling (B.1),

$$\mathbf{B} = \frac{1}{p}\mathbf{X}\mathbf{X}^* = \frac{1}{p}\mathbf{C}\mathbf{Y}\boldsymbol{\Sigma}\mathbf{Y}^*\mathbf{C}^* + \frac{1}{p}\mathbf{C}\mathbf{Y}\boldsymbol{\Sigma}^{1/2}\mathbf{X_0}^* + \frac{1}{p}\mathbf{X_0}\boldsymbol{\Sigma}^{1/2}\mathbf{Y}^*\mathbf{C}^* + \frac{1}{p}\mathbf{X_0}\mathbf{X_0}^*.$$

Assumption A6 implies that

$$\|\frac{1}{p}\mathbf{X_0}\mathbf{X_0}^*\|_2 = O_p(T) \tag{B.8}$$

and that

$$\|\frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*\|_2 = O_p\left(T^{1/2}\|\frac{1}{p}\mathbf{CY\Sigma Y^*C^*}\|_2^{1/2}\right). \tag{B.9}$$

We can write $\frac{(1/p)\mathbf{CY\Sigma Y^*C^*}}{\gamma_1}$ as

$$\frac{(1/p)\mathbf{CY\Sigma Y^*C^*}}{\gamma_1} = \frac{\gamma_{1,l}}{\gamma_1}\frac{(1/p)\mathbf{CY\Sigma Y^*C^*}}{\gamma_{1,l}}$$

$$= \frac{\gamma_{1,l}}{\gamma_1}\frac{(1/p)\mathbf{CY_l\Sigma Y_l^*C^*}}{\gamma_{1,l}} + \frac{\gamma_{1,l}}{\gamma_1}\frac{(1/p)\mathbf{C}(\mathbf{Y\Sigma Y^*} - \mathbf{Y_l\Sigma Y_l}^*)\mathbf{C}^*}{\gamma_{1,l}}. \tag{B.10}$$

From (B.5) we have $\lim_{T\to\infty}\frac{\gamma_{1,l}}{\gamma_1} = 1$. This, together with (B.1), Proposition 4, (B.4), (B.8), (B.9) and Lemma C.4 in the supplementary file, implies (2.11).

We next prove the CLT. In fact we just need to prove

$$\left\|\frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*\right\|_2 = o_p(p^{-1/2}T^2). \tag{B.11}$$

Note that Equation (B.9) implies that $\|\frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*\|_2 = O_p(T^{3/2})$. Remark 2 then follows.

The assumption A7 implies that

$$\|\frac{1}{p}\mathbf{X_0}\mathbf{X_0}^*\|_2 = O_p(T). \tag{B.12}$$

Our aim is to prove (B.11). Note that $rank(\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*) = 1$. Recalling Assumption A7, we can then find

$$\left\|\frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*\right\|_2 = \frac{\sqrt{T}}{p}\sqrt{\sum_{t=1}^{T}\left(\sum_{i=1}^{t}\mathbf{y_i'}\mathbf{\Sigma}^{1/2}\mathbf{x_0}\right)^2}, \tag{B.13}$$

$$\sum_{i=1}^{t}\mathbf{y_i'}\mathbf{\Sigma}^{1/2}\mathbf{x_0} = \sum_{i=1}^{t}\mathbf{y_i'}\mathbf{\Sigma}^{1/2}\sum_{k=0}^{\infty}\tilde{b}_k\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}} + \sum_{i=1}^{t}\mathbf{y_i'}\mathbf{\Sigma}^{1/2}\tilde{b}_{-1}\mathbf{\Sigma_2}^{1/2}\tilde{\mathbf{z}}$$

$$+\sum_{i=1}^{t}\mathbf{y_i'}\mathbf{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2}. \tag{B.14}$$

By (2.1) and a variable change we may write

$$\sum_{i=1}^{t}\mathbf{y_i'} = \sum_{j=1}^{t}\mathbf{z_j'}\left(\sum_{i=j}^{t}b_{i-j}\right) + \sum_{j=-\infty}^{0}\mathbf{z_j'}\left(\sum_{i=1}^{t}b_{i-j}\right). \tag{B.15}$$

Let $(\tilde{c}_{-2,1},\dots,\tilde{c}_{-2,p})' = \tilde{\mathbf{c}}_{-2} = \mathbf{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2}$. Assumptions A3 and A7 imply $\|\tilde{\mathbf{c}}_{-2}\|^2 = O(p)$. Then

$$\sum_{i=1}^{t}\mathbf{y_i}'\mathbf{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2} = \sum_{i=1}^{t}\mathbf{y_i}'\tilde{\mathbf{c}}_{-2}.$$

It follows that

(B.16)
$$E\left(\sum_{i=1}^{t}\mathbf{y_i}'\mathbf{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2}\right) = 0$$

and

(B.17)
$$Var\left(\sum_{i=1}^{t}\mathbf{y_i}'\mathbf{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2}\right) = \|\tilde{\mathbf{c}}_{-2}\|^2\left(\sum_{j=1}^{t}\left(\sum_{i=j}^{t}b_{i-j}\right)^2 + \sum_{j=-\infty}^{0}\left(\sum_{i=1}^{t}b_{i-j}\right)^2\right) = O(pt),$$

which imply

(B.18)
$$\sum_{i=1}^{t}\mathbf{y_i}'\mathbf{\Sigma}^{1/2}\tilde{\mathbf{b}}_{-2} = O_p(p^{1/2}t^{1/2}).$$

As in (B.15), write

$$\sum_{i=1}^{t}\mathbf{y_i}'\mathbf{\Sigma}^{1/2}\tilde{b}_{-1}\mathbf{\Sigma_2}^{1/2}\tilde{\mathbf{z}} = \tilde{b}_{-1}\left(\sum_{j=1}^{t}\mathbf{z_j}'\mathbf{\Sigma}^{1/2}\mathbf{\Sigma_2}^{1/2}\tilde{\mathbf{z}}\left(\sum_{i=j}^{t}b_{i-j}\right) + \sum_{j=-\infty}^{0}\mathbf{z_j}'\mathbf{\Sigma}^{1/2}\mathbf{\Sigma_2}^{1/2}\tilde{\mathbf{z}}\left(\sum_{i=1}^{t}b_{i-j}\right)\right).$$

Assumption A7 implies that $\tilde{\mathbf{z}}$ is independent of $\mathbf{z_t}$ and that $\tilde{b}_{-1}$ is bounded. It follows that

(B.19)
$$\sum_{i=1}^{t}\mathbf{y_i}'\mathbf{\Sigma}^{1/2}\tilde{b}_{-1}\mathbf{\Sigma_2}^{1/2}\tilde{\mathbf{z}} = O_p(p^{1/2}t^{1/2}).$$

Now we consider the first term of the right hand of (B.14). From (B.15), write

$$\begin{aligned}&\sum_{i=1}^{t}\mathbf{y_i}'\mathbf{\Sigma}^{1/2}\sum_{k=0}^{\infty}\tilde{b}_k\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}}\\ &= \sum_{j=1}^{t}\sum_{k=0}^{\infty}\mathbf{z_j}'\mathbf{\Sigma}^{1/2}\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}}\tilde{b}_k\left(\sum_{i=j}^{t}b_{i-j}\right) + \sum_{j=-\infty}^{0}\sum_{k=0}^{\infty}\mathbf{z_j}'\mathbf{\Sigma}^{1/2}\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}}\tilde{b}_k\left(\sum_{i=1}^{t}b_{i-j}\right).\end{aligned}$$

Direct calculations imply

(B.20)
$$E\left(\sum_{i=1}^{t}\mathbf{y_i}'\mathbf{\Sigma}^{1/2}\sum_{k=0}^{\infty}\tilde{b}_k\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}}\right) = \sum_{k=0}^{\infty}\mathrm{tr}\left(\mathbf{\Sigma}^{1/2}\mathbf{\Sigma_1}^{1/2}\right)\tilde{b}_k\left(\sum_{i=1}^{t}b_{i+k}\right) = O(p)$$

and

$$Var\left(\sum_{i=1}^{t}\mathbf{y_i'}\mathbf{\Sigma}^{1/2}\sum_{k=0}^{\infty}\tilde{b}_k\mathbf{\Sigma_1}^{1/2}\mathbf{z_{-k}}\right)=O(pt). \tag{B.21}$$

Equations (B.18)–(B.21) and Assumption A4 imply

$$\|\frac{1}{p}\mathbf{CY\Sigma}^{1/2}\mathbf{X_0}^*\|_2=O_p(\max(p^{-1/2}T^{3/2},T))=o_p(p^{-1/2}T^2). \tag{B.22}$$

The proof of Theorem 2.2 is completed. □

Proof of Theorem 2.1. Define $\mathbf{X_{0\Pi}}=(\mathbf{\Pi x_0},\cdots,\mathbf{\Pi^T x_0})'$ and $\mathbf{X_{1\Pi}}=\mathbf{X}-\mathbf{X_{0\Pi}}$. Write

$$\begin{aligned}\mathbf{B}&=(1/p)\mathbf{XX}^* \\ &=(1/p)\mathbf{X_{1\Pi}X_{1\Pi}}^*+(1/p)\mathbf{X_{1\Pi}X_{0\Pi}}^*+(1/p)\mathbf{X_{0\Pi}X_{1\Pi}}^*+(1/p)\mathbf{X_{0\Pi}X_{0\Pi}}^*.\end{aligned} \tag{B.23}$$

Observe that

$$\|(1/p)\mathbf{X_{0\Pi}}^*\mathbf{X_{0\Pi}}\|_2=\|(1/p)\sum_{t=1}^{T}\mathbf{\Pi}^t\mathbf{x_0x_0'}\mathbf{\Pi'}^t\|_2\le\frac{1}{p(1-\varphi^2)}\|\mathbf{x_0}\|^2. \tag{B.24}$$

This, together with Assumption A6, implies

$$\|(1/p)\mathbf{X_{0\Pi}}^*\mathbf{X_{0\Pi}}\|_2=O_p(1). \tag{B.25}$$

Recalling (C.1), we have

$$\|(1/p)\mathbf{X_{1\Pi}}^*\mathbf{X_{1\Pi}}\|_2\le\frac{M_0}{(1-\varphi)^2}\|(1/p)\mathbf{Y}^*\mathbf{Y}\|_2.$$

We then conclude from (C.1), (C.2) and (B.7) that

$$\lim_{T\to\infty}P\left(\|(1/p)\mathbf{X_{1\Pi}}^*\mathbf{X_{1\Pi}}\|_2\le\frac{8\sum_{i\ge0}|a_i|}{(1-\varphi)^2}M_0\left(1+\sqrt{\frac{T}{p}}\right)^2\right)=1. \tag{B.26}$$

By Holder's inequality

$$\|(1/p)\mathbf{X_{0\Pi}X_{1\Pi}}^*\|_2\le\sqrt{\|(1/p)\mathbf{X_{0\Pi}}^*\mathbf{X_{0\Pi}}\|_2\|(1/p)\mathbf{X_{1\Pi}}^*\mathbf{X_{1\Pi}}\|_2}. \tag{B.27}$$

Thus, equations (B.25)-(B.27) ensure Theorem 2.1.

□

The proof of Theorem 2.3 is simple since $\bar{\mathbf{B}}=\mathbf{HBH}$. So $\|\bar{\mathbf{B}}\|_2\le\|\mathbf{B}\|_2$ since $\|\mathbf{H}\|_2=1$.

Theorem 2.4 are similar to Theorem 2.2. We only need to replace the results of Appendix A.2 by those in Appendix A.3. Note that we don't need to prove (B.11) since (B.2) implies that $\mathbf{x_0}$ doesn't affect $\bar{\mathbf{B}}$.

PROOF OF THEOREM 3.1. At first we prove that the error of the estimator μ_{m_1} is $o(p^{-1/2})$. Let $m_1 = [\sqrt{p}]$. From (2.5) and (B.6) we have

$$\left|\left(a_0 + 2\sum_{1\le j\le m_1} a_j(-1)^j\cos(j\theta_1)\right) - \left(a_0 + 2\sum_{1\le j\le \infty} a_j(-1)^j\cos(j\theta_1)\right)\right|$$

$$\le 2\sum_{1+m_1\le j\le\infty}|a_j| = o(p^{-1/2}) \tag{B.28}$$

and

$$\left|\left(a_0 + 2\sum_{1\le j\le m_1} a_j(-1)^j\cos(j\theta_1)) - (a_0 + 2\sum_{1\le j\le m_1} a_j\right)\right|$$

$$\le 2\sum_{1\le j\le m_1}|a_j|\left(1-\cos\frac{j\pi}{2T+1}\right) = O(p^{1/2}T^{-2}) = o(p^{-1/2}). \tag{B.29}$$

In view of (2.6) it suffices to prove

$$|\mu_{m_1} - (a_0 + 2\sum_{1\le j\le m_1} a_j)\frac{tr(\mathbf{\Sigma})}{p}| = o_p(p^{-1/2}). \tag{B.30}$$

A direct calculation shows the following mean and variance

$$E\mu_{m_1} - \left(a_0 + 2\sum_{1\le j\le m_1} a_j\right)\frac{tr(\mathbf{\Sigma})}{p} = 0,$$

$$Var\left(\sum_{1\le j\le m_1}\frac{1}{T-j-1}\sum_{2\le i\le T-j}\frac{\mathbf{y_i'\Sigma y_{i+j}}}{p}\right)$$

$$= \sum_{1\le i,j\le m_1}\sum_{2\le f\le T-i}\sum_{2\le g\le T-j}\frac{Cov(\frac{\mathbf{y_f'\Sigma y_{f+i}}}{p}, \frac{\mathbf{y_g'\Sigma y_{g+j}}}{p})}{(T-i-1)(T-j-1)}. \tag{B.31}$$

Moreover, we have

$$Cov\left(\frac{\mathbf{y_f'\Sigma y_{f+i}}}{p}, \frac{\mathbf{y_g'\Sigma y_{g+j}}}{p}\right) = \frac{1}{p}\left[\frac{\sum_{i=1}^{p}\Sigma_{ii}^2}{p}E|Z_{ij}|^4\sum_{k=0}^{\infty}b_kb_{k+i}b_{k+g-f}b_{k+g-f+j}1_{(k+g-f\ge 0)}\right.$$

$$\left. + \frac{tr(\mathbf{\Sigma}^2)}{p}E|Z_{ij}|^2(a_{|f-g|}a_{|f+i-g-j|} + a_{|f+i-g|}a_{|f-g-j|})\right].$$

From the above, Assumption (A1) and (B.6), we conclude that

$$Var(\mu_{m_1}) = O(p^{-1}m_1T^{-1}) = o(p^{-1}). \tag{B.32}$$

Then (B.31) and (B.32) imply (B.30).

Now we prove

$$\frac{\sqrt{\frac{|S_{\sigma^2,0,0}|}{p}}}{\sum_{i=2}^{T}\frac{\check{\mathbf{x}}_{i,i}}{p(T-1)}} - \frac{\sqrt{\frac{tr(\mathbf{\Sigma^2})}{p}}}{\frac{tr(\Sigma)}{p}} \xrightarrow{i.p.} 0.$$

Let $\tilde{S}_{\sigma^2,0,0} = S_{\sigma^2,0,0} - a_0^2 tr(\mathbf{\Sigma^2})$. It is then sufficient to show that

$$\frac{\tilde{S}_{\sigma^2,0,0}}{a_0^2 tr(\mathbf{\Sigma^2})} = o_p(1).$$

From Assumptions A2 and A3 we have for large enough T,

$$a_0^2 tr(\mathbf{\Sigma^2}) \geq a_0^2 M_1^2 p, \tag{B.33}$$

where we have used the fact that $\mathrm{tr}(\mathbf{\Sigma^2}) \geq \frac{(tr\mathbf{\Sigma})^2}{p}$. When T is large enough,

$$\left(T - \frac{3}{2}[T/2]\right)([T/2]-1) \geq \frac{T^2}{9}. \tag{B.34}$$

We next expand $\tilde{S}_{\sigma^2,0,0}$ in terms of Z_{ij} and write it a sum of the terms involving the high order of Z_{ij} and the terms involving the low order of Z_{ij}. Specifically, write $\tilde{S}_{\sigma^2,0,0} = \tilde{S}_{\sigma^2,0,0,h} + \tilde{S}_{\sigma^2,0,0,l}$, where

(B.35)

$$\begin{aligned}
&\tilde{S}_{\sigma^2,0,0,h} \\
= \;& \frac{1}{(T - \frac{3}{2}[T/2])([T/2]-1)} \sum_{f=2}^{[T/2]} \sum_{g=f+[T/2]}^{T} \\
&\Big(\sum_{i_1,i_2=1}^{p} \Sigma_{i_1 i_1}\Sigma_{i_1 i_2} \sum_{s_1,s_2=-\infty}^{T} Z_{s_1 i_1}^3 Z_{s_2 i_2} (b_{f-s_1} b_{g-s_1} b_{f+i-s_1} b_{g+j-s_2} \\
&+ b_{f-s_1} b_{g-s_1} b_{f+i-s_2} b_{g+j-s_1} + b_{f-s_1} b_{g-s_2} b_{f+i-s_2} b_{g+j-s_1} \\
&+ b_{f-s_2} b_{g-s_1} b_{f+i-s_1} b_{g+j-s_1}) - 3\sum_{i_1=1}^{p} \Sigma_{i_1 i_1}^2 \sum_{s_1=-\infty}^{T} Z_{s_1 i_1}^4 b_{f-s_1} b_{g-s_1} b_{f+i-s_1} b_{g+j-s_1} \Big).
\end{aligned}$$

Note that $b_k = 0$ when $k < 0$. We can then conclude from Assumption A1 that

$$E|\tilde{S}_{\sigma^2,0,0,h}| = o(p^2 T^{-2}). \tag{B.36}$$

(B.33) and (B.36) imply that

$$\frac{E|\tilde{S}_{\sigma^2,0,0,h}|}{a_0^2 tr(\mathbf{\Sigma^2})} = o(pT^{-2}) = o(1). \tag{B.37}$$

It can be derived that

$$(T-\frac{3}{2}[T/2])([T/2]-1)E\tilde{S}_{\sigma^2,0,0,l} = \sum_{f=2}^{[T/2]}\sum_{g=f+[T/2]}^{T}(a_{g-f}a_{g-f}tr(\mathbf{\Sigma^2})+a_{g-f}a_{g-f}(tr(\mathbf{\Sigma}))^2) = o(p^2T^{-1}). \tag{B.38}$$

This, together with (B.33) and (B.34), implies that

$$\frac{E\tilde{S}_{\sigma^2,0,0,l}}{a_0^2tr(\mathbf{\Sigma^2})} = o(pT^{-3}) = o(1). \tag{B.39}$$

By (B.33), (B.34) and the assumption A1, one can also verify that

$$Var\left(\frac{\tilde{S}_{\sigma^2,0,0,l}}{a_0^2tr(\mathbf{\Sigma^2})}\right) = o(pT^{-2}+p^{-1}) = o(1). \tag{B.40}$$

This, together with (B.37) and (B.39), shows that

$$\frac{\tilde{S}_{\sigma^2,0,0}}{a_0^2tr(\mathbf{\Sigma^2})} = o_p(1).$$

This, together with (B.28)-(B.30), implies that $S_{\sigma^2,m_2} - \frac{a_0\sqrt{2tr(\mathbf{\Sigma^2})}}{\sqrt{p}} \xrightarrow{i.p.} 0$ when m_2 tends to infinity. Since the two estimators are available, it's easy to complete the proof with Theorems 2.2 and 2.4. □

Proof of Theorem 3.2. We claim that

$$\sum_{i=2}^{T}\frac{\check{\mathbf{x}}_{i,i}}{p(T-1)} - \frac{2a_0tr(\mathbf{\Sigma})}{p(1+\varphi)} \xrightarrow{i.p.} 0 \tag{B.41}$$

and

$$S_{\sigma^2,0} - \frac{2a_0\sqrt{2tr(\mathbf{\Sigma^2})}}{\sqrt{p}(1+\varphi)} \xrightarrow{i.p.} 0. \tag{B.42}$$

Indeed, the proofs of (B.41) and (B.42) are similar to that of Theorem 3.1 (replacing $m_1 = m_2$ there by 0). Moreover from Theorem 2.3 we have $\bar{\rho}_1 = o_p(T)$. This, together with (B.41) and (B.42), ensures that

$$\bar{T}_N + \sqrt{\frac{p}{2}}\frac{\frac{tr(\Sigma)}{p}}{\sqrt{\frac{tr(\mathbf{\Sigma^2})}{p}}} \xrightarrow{i.p.} 0, \tag{B.43}$$

which further yields (3.8).

When $\|\phi\|_2 = O(p)$, from Theorem 2.1 we have $\rho_1 = O_p(T)$. This, together with (B.41) and (B.42), ensures that

$$T_N + \sqrt{\frac{p}{2}}\frac{\frac{tr(\Sigma)}{p}}{\sqrt{\frac{tr(\mathbf{\Sigma^2})}{p}}} \xrightarrow{i.p.} 0, \tag{B.44}$$

which further implies (3.9). □

REFERENCES

[1] Bai, Z. D. and Silverstein, J. W. (2006). *Spectral Analysis of Large Dimensional Random Matrices*. 2nd Edition, Springer, New York.

[2] 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.

[3] 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.

[4] Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivar. Anal.* **97**, 1382–1408.

[5] 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.

[6] Chan, N.H. and Wei, C.Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367–401.

[7] Chang, Y. (2004). Bootstrap unit root tests in panels with cross sectional dependency. *J. Econom.* **120**, 263–293.

[8] 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.

[9] Choi, In. (2001). Unit root tests for panel data. *J. Internat. Money & Finan.* **20**, 249–272.

[10] Choi, In. (2015). *Almost All about Unit Roots: Foundations, Developments and Applications*. Cambridge University Press, London.

[11] Choi, In. and Chue, T. K. (2007). Subsampling hypothesis tests for nonstationary panels with applications to exchange rates and stock prices. *J. Appl. Econom.* **22**, 233–264.

[12] 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.

[13] 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.

[14] Fisher, R. A. (1937). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugenics.* **9**, 238–249.

[15] Hsu, P. L. (1939). On the distribution of roots of certain determinant equations, *Ann. Eugenics.* **9**, 250–258.

[16] Im, K., Pesaran, M.H. and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *J. Econom.* **115**, 53–74.

[17] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.* **29**, 295–327.

[18] Johnstone, I.M. (2007). High dimensional statistical inference and random matrices. *In International Congress of Mathematicians I*, 307–333.

[19] Levin, A., Lin, C.F. and Chu, C.S.J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *J. Econom.* **108**, 1–24.

[20] Liu, H. Y., Aue. A. and Paul. D. (2015). On the Marcenko–Pastur law for linear time series. *Ann. Statist.* **43**, 675–712.

[21] Maddala, G. S. and Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford B. Econ. & Statist.* **61**, 631–652.

[22] Marčenko, V. A. and Pastur, L. A. (1967). Distribution for some sets of random matrices, *Math. USSR-Sb.* **1**, 457–483.

[23] Moon, H.R. Perron, B and Phillips, P.C.B. (2007). Incidental trends and the power of panel unit root tests. *Journal of Econometrics.* **141**, Issue 2, 416–459.

[24] Pan, G., Gao, J and Yang, Y. (2014). Testing independence among a large number of high dimensional random vectors. *J. Amer. Statist. Assoc.* **109**, 600–612.

[25] Paul, D. (2007). Asymptotics of sample eigenvalue structure for a large dimensional spiked covariance model. *Statist. Sinica.* **17**, 1617–1642.

[26] Paul, D. and Aue, A. (2014). Random matrix theory in statistics: a review. *J. Statist. Plan. & Infer.* **150**, 1–29.

[27] Pesaran, M.H. (2007). A simple panel unit root test in the presence of cross–sectional dependence. *J. Appl. Econom.* **22**, 265–312.

[28] Pesaran, M.H. (2015). *Time Series and Panel Data Econometrics.* Oxford University Press, Oxford.

[29] Pesaran, M.H., Smith, L. V. and Yamagata, T. (2013). Panel unit root tests in the presence of a multifactor error structure. *J. Econom.* **175**, 94–115.

[30] Phillips, P and Perron, P (1988). Testing for a unit root in time series regression. *Biometrika* **75**, 335–346.

[31] Roy, S. N. (1939). *P*–statistics and some generalizations in analysis of variance appropriate to multivariate problems. *Sankhyā.* **4**, 381–396.

[32] Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108**, 1033–1056.

[33] Tracy, C. A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159**, 151–174.

[34] Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177**, 727–754.

[35] Yao, J.F., Zheng, S. R. and Bai, Z.D. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press, Cambridge.

[36] Zhang, L. (2006). Spectral analysis of large dimensional random matrices. Ph.D. thesis, National University of Singapore, Singapore.

Bo Zhang and Guangming Pan
School of Physical and Mathematical Sciences
Nanyang Technological University, Singapore
bo.zhang1@monash.edu
gmpan@ntu.edu.sg

Bo Zhang and Jiti Gao
Department of Econometrics
and Business Statistics
Monash University, Australia
bo.zhang1@monash.edu
jiti.gao@monash.edu