ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Testing Indepedence for a Large Number of High-Dimensional Random Vectors

Guangming Pan, Jiti Gao and Yanrong Yang

March 2013

Working Paper 09/13

Testing Independence for a Large Number of High–Dimensional Random Vectors

Guangming Pan, Jiti Gao and Yanrong Yang

March 18, 2013

Abstract

Capturing dependence among a large number of high dimensional random vectors is a very important and challenging problem. By arranging n random vectors of length p in the form of a matrix, we develop a linear spectral statistic of the constructed matrix to test whether the n random vectors are independent or not. Specifically, the proposed statistic can also be applied to n random vectors, each of whose elements can be written as either a linear stationary process or a linear combination of a random vector with independent elements. The asymptotic distribution of the proposed test statistic is established in the case of $0 < \lim_{n\to\infty} \frac{p}{n} < \infty$ as $n \to \infty$. In order to avoid estimating the spectrum of each random vector, a modified test statistic, which is based on splitting the original n vectors into two equal parts and eliminating the term that contains the inner structure of each random vector or time series, is constructed. The facts that the limiting distribution is a normal distribution and there is no need to know the inner structure of each investigated random vector result in simple implementation of the constructed test statistic. Simulation results demonstrate that the proposed test is powerful against many common dependent cases. An empirical application to detecting dependence of the closed prices from several stocks in S&P500 also illustrates the applicability and effectiveness of our provided test.

Keywords: Central limit theorem, Covariance stationary time series, Empirical spectral distribution, Independence test, Large dimensional sample covariance matrix; Linear spectral statistics.

JEL Classifications: C12, C21, C22.

1 Introduction

Testing cross-sectional dependence between a large number of high–dimensional random vectors attracts great interest in high dimensional statistical analysis, especially in longitudinal data and panel data analysis (Frees (1995); Mundlak (1978); Hsiao, Pesaran and Pick (2009); Sarafidis, Yamagata and Robertson (2009); Chen, Gao and Li (2012)). In longitudinal data or panel data analysis, one of the key reasons of pooling the data together is to overcome the aggregation problems that arise with dependent data in modelling the behaviour of heterogenous agents on the basis of the representative assumption. In multivariate time series analysis, elucidation of various causalities between time series is vital to forecasting and prediction. Compared with the literature focusing on detecting serial dependence within a univariate time series, relatively few studies have been done to capture dependence between time series (Haugh (1976); Geweke (1981); Hong (1996)). Moreover, the goal of these papers is restricted to investigating dependence between two covariance stationary time series.

Mutual independence is difficult to test while nonlinear dependence is also not easy to detect. Mutual independence is more demanding than pairwise independence. One conventional measure of linear dependence is the correlation function, which may overlook nonlinear dependent structures that have zero correlations, e.g. Hong (1996). Another useful tool is to utilize the equivalence of the joint distribution and the product of the corresponding marginal distributions under independent case (see Hong (2000); Hong (2005)). Of course, this method can capture all kinds of dependence types since it is a sufficient and necessary condition of independence. However, it is just applicable to pairwise independence test rather than mutual independence test for a large number of high–dimensional random vectors. Hong (1999) developed a generalized spectral density approach via the empirical characteristic function for serial independence test of one time series. This method is also applicable to some types of linear and nonlinear dependencies but only works for detecting pairwise dependence.

In this paper, we propose a novel test statistic to test mutual independence for n random vectors of length p when n and p are comparable. Since there are $n \times p$ observed data available, we pool them together to form a data matrix so that some features of the data matrix to investigate independence among the initial n random vectors can be utilized. Large dimensional random matrix theory then serves as a powerful tool to investigate such a matrix. Specifically speaking, we group the n random vectors into a matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ and then consider the empirical spectral distribution (ESD) of the eigenvalues of the corresponding sample covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$, where $\mathbf{x}_i, i = 1, 2 \ldots, n$ are the observed n time series, each being of length p, i.e. $\mathbf{x}_i = (X_{1i}, X_{2i}, \ldots, X_{pi})'$. Here we would like to point out that there have been a substantial set of research works dealing with high dimensional data by random matrix theory (see, for example, Ledoit and Wolf (2002), Johnstone (2001), Birke and Dette (2005) and Yao (2012)). Our approach essentially uses the ESD of the sample covariance matrix $\mathbf{S}$ for n random vectors to distinguish dependence from independence. Our discussion covers both the case where the random vectors are n covariance stationary time series and the case where the random vectors are vectors of linear combinations of independent random variables.

To study the size of the proposed test we first establish the limiting spectral distribution(LSD), i.e. the limit of the ESD of the sample covariance matrix $\mathbf{S}$ under the finite second moment condition on the components. This generalizes the result of Yao (2012), which obtained the LSD under the finite fourth moment condition. Moreover, for the first time we establish a central limit theorem (CLT) for linear spectral statistics of the sample covariance matrices whose columns are covariance stationary time series under the finite fourth moment condition on the time series components. This CLT complements the classical result of linear spectral statistics of the sample covariance matrices of the independent random vectors with i.i.d. components or linear independent structure (see Bai and Silverstein (2009) and Lytova and Pastur (2009)).

As stated above, correlation functions are useful enough for describing linear dependence but can not detect all sorts of nonlinear dependecies. To some extent, our proposed test statistic is also based on a correlation structure, i.e. the sample covariance matrix. A natural question is how our test performs under all sorts of dependent structures. For the Gaussian case, the sample covariance matrix of a linear covariance stationary time series can be written in the form of $\mathbf{S}_1 = \frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{Y}\mathbf{Y}'\mathbf{T}_1^{1/2}$, where $\mathbf{T}_1$ is a $p \times p$ nonnegative positive Hermitian deterministic matrix and $\mathbf{Y}$ is a $p \times n$ random matrix with i.i.d. components. If the cross–sectional dependence can be described as $\frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{Y}\mathbf{T}_2\mathbf{Y}'\mathbf{T}_1^{1/2}$ with $\mathbf{T}_2$ being an $n \times n$ Hermitian deterministic matrix, the limit of its ESD is then given in Theorem 1.2.1 of Zhang (2006), which is different from the limit of the ESD of $\mathbf{S}_1$ corresponding to the independent case. In view of this, our test is able to capture this type of dependent structure. In panel data analysis, the issue of how to characterise cross–sectional dependence attracts great attention among researchers. Spatial models and factor models are two commonly used dependent structures. The simulation given in Section 4 below shows that the proposed test can be applied to these two types of dependence. Finite sample simulations illustrate that the proposed test can also detect some kinds of nonlinear dependence with zero correlations except the "ARCH" dependence. To detect the ARCH dependence we use high power of entries X_{ji} instead of X_{ji} so that the test statistic still works.

The paper is organized as follows. In Section 2, we briefly review some basic concepts and results from large dimensional random matrix theory. Section 3 states the proposed test statistic and the asymptotic theorems for the developed statistic, including the LSD of the sample covariance matrix for n covariance stationary time series and the CLT of the linear spectral statistic. Section 4 analyses the finite sample performance of the test and investigate some kinds of commonly used cross–sectional dependent structures, including non-zero correlation dependences (e.g. spatial models and factor models, etc.) and some zero-correlation dependent structures. Section 5 provides an empirical application to stock prices in S&P 500. Section 6 presents a conclusion. All the proofs are given in an appendix. Throughout the paper, the limit is taken as $n \to \infty$.

2 Preliminaries

The observed n random vectors $\mathbf{x}_i = (X_{1i}, X_{2i}, \ldots, X_{pi})'$ with $i = 1, 2, \ldots, n$ are grouped into a matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$. Denote the sample covariance matrix by

$$\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T. \tag{2.1}$$

The goal is to do the following independence hypothesis test

$$\mathbb{H}_0 : \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \textit{ are independent}; \ \textit{against } \mathbb{H}_1 : \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \textit{ are dependent}.$$

Throughout the paper, we consider two types of high dimensional random vectors $\mathbf{x}_i$. The first type $\mathbf{x}_i$ is stationary time series specified as follows.

Assumption 1. *The n time series can be expressed as*

$$X_{jt} = \sum_{k=0}^{\infty} b_k \xi_{j-k,t}, \ j = 1, \ldots, p; \ t = 1, \ldots, n, \tag{2.2}$$

where for any $t = 1, 2, \dots, n$, $\{\xi_{i,t}\}_{i=-\infty}^{\infty}$ is an independent and identically distributed (i.i.d) sequence with mean zero and variance one; $\{b_k\}_{k=0}^{\infty}$ is a sequence of real numbers satisfying $\sum_{k=0}^{\infty} |b_k| < \infty$.

This assumption covers many classical covariance stationary time series, for example, the autoregressive (AR), moving average (MA), and autoregressive and moving average(ARMA) time series of finite orders, etc.. In addition to ensuring stationary, the condition $\sum_{k=0}^{\infty} |b_k| < \infty$ is imposed to also guarantee that the spectral norm of the population covariance matrix $\mathbf{T}_1$ of each time series under investigation is bounded, as will be seen from the proof.

The second type $\mathbf{x}_i$ is linearly generated by $\mathbf{y}_i$ whose components are independent, as defined below.

Assumption 2. *Let $\mathbf{x}_i = \mathbf{T}_1^{1/2}\mathbf{y}_i$ with $\mathbf{y}_i = (Y_{1i}, \cdots, Y_{pi})^T$ and $\mathbf{T}_1^{1/2}$ being a Hermitian square root of the nonrandom nonnegative definite Hermitian matrix $\mathbf{T}_1$. For each $i = 1, \dots, n$, $Y_{1i}, \cdots, Y_{pi}$ are i.i.d with mean zero and variance one.*

Assumption 3. *Let p be some function of n. Assume that n and p tend to infinity in the same order, i.e.*

$$c := \lim_{n\to\infty} \frac{p}{n} \in (0, +\infty).$$

When $\{\xi_{i,t}\}$ are normally distributed, Assumption 1 is a special case of Assumption 2. Indeed, it is clear that if $\{\xi_{i,t}\}$ are normally distributed, each X_{jt} has a Gaussian distribution and each $\mathbf{x}_i$ has a multivariate Gaussian distribution whose covariance matrix is a Toeplitz matrix. Then $\mathbf{x}_i$ in Assumption 1 can be written as a form of $\mathbf{T}_1^{1/2}\mathbf{y}_i$ as well. Here, to save notation, we still use $\mathbf{T}_1$ as a covariance matrix of $\mathbf{x}_i$ although it is a Toeplitz matrix. Therefore in this case the sample covariance matrices $\mathbf{S}$ associated with Assumptions 1 and 2 have a unified expression

$$\frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{Y}\mathbf{Y}^T\mathbf{T}_1^{1/2}, \tag{2.3}$$

where $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$.

Denote the sample covariance matrix in the form of (2.3) by $\mathbf{S}_1$. We are now interested in its limiting spectral distribution (LSD) which is the limit of the empirical spectral distribution(ESD) $F^{\mathbf{S}_1}(x)$. Here for any $\mathbf{A}$ of size $p \times p$ with real eigenvalues, its ESD is defined by

$$F^{\mathbf{A}}(x) = \frac{1}{p}\sum_{j=1}^{p} I(\mu_j \le x),$$

where $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_p$ are eigenvalues of the matrix $\mathbf{A}$. A common way to find the LSD is to first establish an equation of its Stieltjes transform, which is defined as, for any cumulative distribution function (CDF) $G(x)$,

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \; Im(z) \neq 0.$$

It can be then recovered by the Frobenius-Perron formula inversion formula

$$G\{[a, b]\} = \frac{1}{\pi} \lim_{\eta \to 0^+} \int_a^b Im\Big(m_G(\zeta + i\eta)\Big) d\zeta, \tag{2.4}$$

where a, b are points of continuity of $G(x)$.

Silverstein's result (1995) shows that the LSD of $\mathbf{S}_1$ in (2.3) is $F_{c,H}(x)$ whose Stieltjes transform is the unique solution to

$$m(z) = \int \frac{1}{\tau\Big(1 - c - czm(z)\Big) - z} dH(\tau), \tag{2.5}$$

in the set $\{m \in \mathbb{C} : -\frac{1-c}{z} + cm \in \mathbb{C}^+\}$ if $F^{\mathbf{T}_1} \to H(\tau)$. This also yields the LSD of the sample covariance matrix $\mathbf{S}$ for linear stationary processes with the Gaussian entries because the condition that $F^{\mathbf{T}_1} \to H(\tau)$ holds automatically in the case of linear stationary time series. An alternative expression of (2.5) for stationary time series will be given in the next section by using its spectral density.

To propose a statistic to test the hypothesis $\mathbb{H}_0$ based on the feature of $F_{c,H}(x)$, we consider an alternative that the sample covariance matrix $\mathbf{S}$ takes the form of

$$\frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{Y}\mathbf{T}_2\mathbf{Y}^T\mathbf{T}_1^{1/2}, \tag{2.6}$$

where $\mathbf{T}_2$ is an $n \times n$ deterministic Hermitian matrix. Hence the dependence of the n time series is described by the matrix $\mathbf{T}_2$.

Denote the sample covariance matrix in the form of (2.6) by $\mathbf{S}_2$. Zhang (2006) provides the LSD of the matrix $\mathbf{S}_2$ different from (2.5). For easy reference, we state this result in the following lemma.

Lemma 1. *In addition to Assumptions 2 and 3, we assume that as $n \to \infty$, the ESDs of $\mathbf{T}_1$ and $\mathbf{T}_2$, denoted by $F^{\mathbf{T}_1}$ and $F^{\mathbf{T}_2}$ respectively, converge weakly to two probability functions, H_1 and H_2, respectively. Then the ESD of the matrix $\mathbf{S}_2$ converges weakly to a* *non-random CDF* F_{c,H_1,H_2} *with probability one, for which if* $H_1 \equiv 1_{[0,+\infty)}$ *or* $H_2 \equiv 1_{[0,+\infty)}$, *then* $F_{c,H_1,H_2} \equiv 1_{[0,+\infty)}$*; otherwise if for each* $z \in \mathbb{C}^+$,

$$\begin{cases} s(z) = -z^{-1}(1-c) - z^{-1}c \int \frac{1}{1+q(z)x} dH_2(x) \\ s(z) = -z^{-1} \int \frac{1}{1+p(z)y} dH_1(y) \\ s(z) = -z^{-1} - p(z)q(z) \end{cases} \tag{2.7}$$

is viewed as a system of equations for the complex vector $(s(z), p(z), q(z))$*, then the Stieltjes transform of* F_{c,H_1,H_2}*, denoted by* $m_{F_{c,H_1,H_2}}(z)$*, together with two other functions, denoted by* $g_1(z)$ *and* $g_2(z)$*, both of which are analytic on* $\mathbb{C}^+$*, will satisfy that* $\big(m_{F_{c,H_1,H_2}}(z), g_1(z), g_2(z)\big)$ *is the unique solution to (2.7) in the set*

$$\Big\{\big(s(z), p(z), q(z)\big) : Im\big(s(z)\big) > 0, Im\big(zp(z)\big) > 0, Im\big(q(z)\big) > 0\Big\}.$$

From (2.5) and (2.7), we see that the LSD of the matrix $\mathbf{S}_1$ is different from that of $\mathbf{S}_2$ since the latter one depends on the spectral distribution of the matrix $\mathbf{T}_2$ which is an identity matrix under the null hypothesis $\mathbb{H}_0$. Based on the observation, a natural idea is to utilize the difference between the LSDs of $\mathbf{S}$ under $\mathbb{H}_0$ and $\mathbb{H}_1$ to distinguish independence from dependence.

To this end let

$$G_n(\lambda) = p\Big(F^{\mathbf{S}}(\lambda) - F_{c_n,H_n}(\lambda)\Big) \tag{2.8}$$

and consider the linear spectral statistic of $\mathbf{S}$:

$$M_n = \int f(\lambda) dG_n(\lambda), \tag{2.9}$$

where $F_{c_n,H_n}(\lambda)$ is obtained from the LSD $F_{c,H}(\lambda)$ of $\mathbf{S}$ under $\mathbf{H}_0$ and Assumptions 1 or 2 with c and H replaced by $c_n = p/n$ and H_n respectively; $H_n = F^{\mathbf{T}_1}$ and $f(\lambda)$ is a smooth function. Roughly speaking, the difference between the LSDs of $\mathbf{S}$ under $\mathbb{H}_0$ and $\mathbb{H}_1$ is reflected in behaviour of M_n. Indeed, if we rewrite the statistic M_n as

$$p\Big[\int f(\lambda) d\Big(F^{\mathbf{S}}(\lambda) - F_{c_n,H_n,\mathbb{H}_1}(\lambda)\Big)\Big] + p\Big[\int f(\lambda) d\Big(F_{c_n,H_n,\mathbb{H}_1}(\lambda) - F_{c_n,H_n}(\lambda)\Big)\Big], \tag{2.10}$$

where $F_{c_n,H_n,\mathbb{H}_1}(\lambda)$ denotes the LSD of $\mathbf{S}$ under the alternative hypothesis $\mathbb{H}_1$, then we see that the last term of (2.10) captures the difference between the LSDs of $\mathbf{S}$ under $\mathbb{H}_0$ and $\mathbb{H}_1$, not to mention the first term of (2.10). One typical example of $F_{c_n,H_n,\mathbb{H}_1}(\lambda)$ could be F_{c,H_1,H_2} in Lemma 1.

Central limit theorems (CLT) of M_n corresponding to Assumptions 1 and 2 will be given in the next section. Based on it we then propose our test statistic.

3 Main theorems and the test statistic

3.1 Covariance stationary time series

The aim of this subsection is to establish the LSD of $\mathbf{S}$ and CLT of the linear spectral statistic M_n under the null hypothesis $\mathbb{H}_0$ and Assumption 1. Below we first present the LSD of $\mathbf{S}$.

Theorem 1. *Under Assumptions 1 and 3 and the null hypothesis $\mathbb{H}_0$, with probability one, the ESD $F^{\mathbf{S}}(x)$ converges to a nonrandom distribution function $F_{c,\phi}(x)$ whose Stieltjes transform $m_\phi(z)$ satisfies*

$$z = -\frac{1}{m_\phi(z)} + \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{cm_\phi(z) + \big(\phi(\lambda)\big)^{-1}} d\lambda, \tag{3.1}$$

where $\phi(\lambda)$ denotes the spectral density of $\mathbf{x}_t$

$$\phi(\lambda) = \sum_{k=-\infty}^{\infty} \phi_k e^{ik\lambda}, \;\; \lambda \in [0, 2\pi),$$

with $\phi_k = Cov(X_{jt}, X_{j+k,t})$.

Remark 1. *This weakens the finite fourth moment condition imposed in Yao (2012). In addition we would point out that (3.1) is just an alternative expression of (2.5) in terms of the spectral density of $\mathbf{x}_i$. Therefore we use $F_{c,\phi}(x)$ to denote $F_{c,H}(x)$ in the case of stationary time series.*

From (3.1), we see that the Stieltjes transform $m_\phi(z)$ does not have an explicit expression. In practice, we can adopt a numerical method to calculate it, which is provided in Yao (2012). For easy reference, we state it below:

Algorithm of calculating $m_\phi(z)$: Choose an initial value $m_\phi^{(0)}(z) = u + i\varepsilon$, where $z = x + i\varepsilon$ with x a given value and ε a small enough number. Iterate the following mapping below for $k \geq 0$:

$$\frac{1}{m_\phi(z)} = -z + A(m_\phi(z)), \tag{3.2}$$

where

$$A(m_\phi(z)) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{cm_\phi(z) + \phi^{-1}(\lambda)} d\lambda,$$

until convergence. Let $m_\phi^{(K)}(z)$ be the final value.

We next develop CLT of M_n, which, we believe, is new in the literature. Recall the definition of $G_n(\lambda)$ in (2.8).

Theorem 2. *In addition to Assumptions 1 and 3, we suppose that* $E\xi_{j-k,t}^4 = 3$. *Let* $f_1, f_2, \ldots, f_h$ *be functions analytic on an open region containing the support of* F_{c_n,H_n}. *Then the random vector*

$$\Big(\int f_1(\lambda)dG_n(\lambda), \int f_2(\lambda)dG_n(\lambda), \ldots, \int f_h(\lambda)dG_n(\lambda)\Big) \tag{3.3}$$

converges in distribution to a Gaussian random vector $\Big(X_{f_1}, X_{f_2}, \ldots, X_{f_h}\Big)$ *with mean function for* $\ell = 1, 2, \cdots, h$,

$$EX_{f_\ell} = -\frac{1}{2\pi i}\oint_{\mathcal{C}} f_\ell(z)\frac{\frac{1}{2\pi}\int_0^{2\pi} c\underline{m}_\phi^3(z)\phi^2(\lambda)\big(1+\phi(\lambda)\underline{m}_\phi(z)\big)^{-3}d\lambda}{\big(1 - c\frac{1}{2\pi}\int_0^{2\pi}\underline{m}_\phi^2(z)\phi^2(\lambda)\big(1+\phi(\lambda)\underline{m}_\phi(z)\big)^{-2}d\lambda\big)^2}dz$$

and covariance element for $\ell, r = 1, 2, \cdots, h$,

$$Cov(X_{f_\ell}, X_{f_r}) = -\frac{1}{2\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2}\frac{f_\ell(z_1)f_r(z_2)}{\big(\underline{m}_\phi(z_1) - \underline{m}_\phi(z_2)\big)^2}\frac{d\underline{m}_\phi(z_1)}{dz_1}\frac{d\underline{m}_\phi(z_2)}{dz_2}dz_1dz_2. \tag{3.4}$$

The contours $\mathcal{C}$ *above are closed and are taken in the positive direction in the complex plane, each enclosing the support of* $F_{c,\phi}(\lambda)$ *and* $\underline{m}_\phi(z)$ *is the Stieltjes transform of the LSD of the matrix* $\underline{\mathbf{S}} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$.

Here $\underline{m}_\phi(z)$ can be obtained from $m_\phi(z)$ of (3.1) because the spectra of $\underline{\mathbf{S}}$ differs from that of $\mathbf{S}$ by $|n-p|$ zeros.

3.2 Linear independent structures

This subsection is to consider $\mathbf{x}_i$ satisfying Assumption 2.

The CLT of the linear spectral statistic M_n defined in (2.9) has been reported in Theorem 9.10 of Bai and Silverstein (2009). For easy reference, we list it below.

Proposition 1. *In addition to Assumptions 2 and 3 suppose that* $EY_{11}^4 = 3$ *and* $\|\mathbf{T}_1\|$, *the spectral norm of* $\mathbf{T}_1$, *is bounded and* $F^{\mathbf{T}_1}$ *converges weakly to* $H(x)$. *Then the random vector (3.3) converges in distribution to a Gaussian vector with mean*

$$EX_f = -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{c\int\frac{\underline{m}^3(z)t^2dH(t)}{(1+t\underline{m}(z))^3}}{\Big(1 - c\int\frac{\underline{m}^2(z)t^2dH(t)}{(1+t\underline{m}(z))^2}\Big)^2}dz \tag{3.5}$$

and covariance function being the same as (3.4) with $\underline{m}_\phi(z)$ *replaced by* $\underline{m}(z)$. *Here* $\underline{m}(z)$, *which can be obtained from* $m(z)$ *in (2.5), is the Stieltjes transform of the LSD of the matrix* $\underline{\mathbf{S}} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$.

When $\mathbf{T}_1$ becomes the identity matrix, $H(t)$ becomes a degenerate distribution at point 1 and we do not need to assume that $EY_{11}^4 = 3$ in this case. Theorem 1.4 of Pan and Zhou (2008) gives CLT for the random vector (3.3). We list it below.

Proposition 2. *In addition to Assumptions 2 and 3 suppose that $EY_{11}^4 < \infty$. Then the random vector (3.3) converges in distribution to a Gaussian vector with mean*

$$EX_f = -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{c\frac{\underline{m}^3(z)}{(1+\underline{m}(z))^3}}{\left(1-c\frac{\underline{m}^2(z)}{(1+t\underline{m}(z))^2}\right)^2}dz - \frac{c(EX_{11}^4-3)}{2\pi^2}\oint_{\mathcal{C}} f(z)\frac{\frac{\underline{m}^3(z)}{(1+\underline{m}(z))^3}}{1-c\frac{\underline{m}^2(z)}{(1+t\underline{m}(z))}}dz \qquad (3.6)$$

and covariance

$$Cov(X_{f_l}, X_{f_r}) = -\frac{1}{\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2}\frac{f_l(z_1)f_r(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\frac{d}{dz_2}\underline{m}(z_2)\frac{d}{dz_1}\underline{m}(z_1)dz_1dz_2$$
$$-\frac{c(EX_{11}^4-3)}{2\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f_l(z_1)f_r(z_2)\frac{d}{dz_1}[\frac{1}{1+\underline{m}(z_1)}]\frac{d}{dz_2}[\frac{1}{1+\underline{m}(z_2)}]dz_1dz_2. \qquad (3.7)$$

3.3 Test statistic

There are two questions to be addressed before proposing a test statistic based on Theorem 2, Propositions 1 and 2. The first one is the choice of the test function $f(\lambda)$ associated with M_n in (2.9). The second one is that the mean of the asymptotic distribution of M_n, which includes the spectral density $\phi(\lambda)$ of time series $\mathbf{x}_i$ or $H(x)$ associated with linear independence structures, is often unknown in practice no matter what $f(\lambda)$ is.

For the first question, we choose two simple test functions $f_1(\lambda) = \lambda$ and $f_2(\lambda) = \lambda^2$ for simplicity and consider their linear combination. To overcome the second difficulty, we divide n time series into two groups, each of which contains $[n/2]$ time series, where $[n/2]$ is the largest integer smaller than $n/2$. By Theorem 2 or Proposition 1 we have

$$\left(\int xdG_n^{(i)}(x), \int x^2dG_n^{(i)}(x)\right) \xrightarrow{d} \left(X_x^{(i)}, X_{x^2}^{(i)}\right),\ as\ n\to\infty,\ i=1,2, \qquad (3.8)$$

where $G_n^{(i)}(x) = p\left(F^{\mathbf{S}^{(i)}}(x) - F_{c_{n^{(i)}},H_{n^{(i)}}}(x)\right)$ with $c_{n^{(i)}} = p/[n/2], H_{n^{(i)}} = H_n$, $F_{c_{n^{(i)}},H_{n^{(i)}}}(x)$ is the analogue of F_{c_n,H_n} but corresponding to $\mathbf{S}^{(i)} = \frac{1}{[n/2]}\mathbf{X}^{(i)}\mathbf{X}^{(i)\prime}$ and $\mathbf{X}^{(i)}$ consisting of the i-th group of the divided time series, $i = 1, 2$ ($\mathbf{X} = (\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ when n is even). Here $\left(X_x^{(i)}, X_{x^2}^{(i)}\right)$ is the limiting distribution corresponding to the i-th group time series. Since the statistics on the left hand side of (3.8) for the two groups of time series are independent under $\mathbf{H}_0$, we calculate the difference of the two statistics and obtain

$$\left(\int xd\widetilde{G}_n(x), \int x^2d\widetilde{G}_n(x)\right) \xrightarrow{d} \left(\widetilde{X}_x, \widetilde{X}_{x^2}\right),\ as\ n\to\infty, \qquad (3.9)$$

where

$$\widetilde{G}_n(x) = G_n^{(1)}(x) - G_n^{(2)}(x) = p\Big(F^{\mathbf{S}^{(1)}}(x) - F^{\mathbf{S}^{(2)}}(x)\Big), \tag{3.10}$$

and $\widetilde{X}_x = X_x^{(1)} - X_x^{(2)}$, $\widetilde{X}_{x^2} = X_{x^2}^{(1)} - X_{x^2}^{(2)}$.

It follows from Theorem 2 that $(\widetilde{X}_x, \widetilde{X}_{x^2})$ is bivariate normal with mean 0 and covariance matrix $\widetilde{\mathbf{\Omega}}$, where $\widetilde{\mathbf{\Omega}} = 2\mathbf{\Omega}$ and $\mathbf{\Omega} = (\omega_{gh})_{2\times 2}$ is the asymptotic covariance matrix of $(\int x dG_n^{(i)}(x), \int x^2 dG_n^{(i)}(x))$ given by

$$\omega_{gh} = -\frac{1}{\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} \frac{f_g(z_1) f_h(z_2)}{(\underline{m}_\phi(z_1) - \underline{m}_\phi(z_2))^2} \frac{d}{dz_2}\underline{m}_\phi(z_2)\frac{d}{dz_1}\underline{m}_\phi(z_1) dz_1 dz_2. \tag{3.11}$$

Note that (3.10) does not involve any unknown parameters. Therefore, we propose the following testing statistic for $\mathbb{H}_0$:

$$L_n = \Big(\int x d\widetilde{G}_n(x), \int x^2 d\widetilde{G}_n(x)\Big)\widetilde{\mathbf{\Omega}}^{-1}\begin{pmatrix} \int x d\widetilde{G}_n(x) \\ \int x^2 d\widetilde{G}_n(x)\end{pmatrix}. \tag{3.12}$$

As for the linear independence structures, the statistic L_n is the same except that $\underline{m}_\phi(z)$ in (3.11) should be replaced by the Stieltjes transform $\underline{m}(z)$ given in Proposition 1.

The following theorem is a direct application of Theorem 2 or Proposition 1.

Theorem 3. *Under the assumptions in Theorem 2 or in Proposition 1, the test statistic L_n converges in distribution to $\chi^2(2)$, which denotes the chi–squared random variable with the degree of freedom being 2.*

Remark 2. *The proposed statistic L_n contains the inverse covariance matrix $\widetilde{\mathbf{\Omega}}^{-1}$ and this matrix contains the unknown parameter $\underline{m}_\phi(z)$. This parameter can be estimated either by the algorithm provided above, or the sample Stieltjes transform $\underline{m}_n(z) = \frac{1}{p}tr(\mathbf{X}'\mathbf{X} - z\mathbf{I}_n)^{-1}$. Furthermore, the asymptotic distribution is still χ^2 after plugging in the estimator of $\underline{m}_\phi(z)$ by the Slutsky theorem. In view of this the proposed statistic L_n is easy to implement.*

Remark 3. *Traditionally, the method of dividing total samples into two parts is to use one part to do test and the other part to estimate unknown parameters. However, the strategy of dividing total samples into two parts here serves as a different purpose, eliminating the unknown term involved in the linear spectral statistic M_n. Indeed, we make use of the full strength of all observations, because if the first group and the second group are not independent or there is dependence among each group, then (3.9) is not true.*

We also considered a Bootstrap method as follows. By a parametric bootstrap we may redraw a sample $\mathbf{x}^* = \{\mathbf{x}_1^*, ..., \mathbf{x}_n^*\}$ *from the p-variate normal distribution with mean zero and the population covariance matrix* $\mathbf{S}$ *defined in (2.1). Then consider the bootstrap linear spectral statistic*

$$\int f(x)dG_n^*(x), \tag{3.13}$$

where $G_n^*(x) = p\Big[F^{\mathbf{S}_3}(x) - F_{c_n,F^{\mathbf{S}}}(x)\Big]$ *and* $\mathbf{S}_3 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^*(\mathbf{x}_i^*)^T$*. We can further construct a statistic like (3.12) by replacing* $\widetilde{G}_n(x)$ *with* $G_n^*(x)$*. Moreover its asymptotic distribution can be directly obtained from Theorem 2 or Proposition 1.*

However simulations show that the bootstrap statistic is not as powerful as the one proposed based on the strategy of dividing observations. The key reason is that the independence assumption under $\mathbb{H}_0$ *is reflected in* $F^{\mathbf{S}}$ *and its limit only such that the difference* $p(F^{\mathbf{S}} - F_{c_n,H_n})$ *is not used. As a consequence it can not identify the alternatives whose limit is the same as the one determined by* (2.6) *such as* $\frac{1}{n}\mathbf{X}\mathbf{T}_3\mathbf{X}^T$ *with* $T_3 = \mathbf{I} + \mathbf{e}\mathbf{e}^T$ *(all components of* $\mathbf{e}$ *are one).*

3.4 The power under local alternatives

This section is to investigate the power for some local alternatives. The first interesting example (local alternative) is that $\mathbf{x}_1, \cdots, \mathbf{x}_n$ satisfy Assumption 2 but $\mathbf{T}_1$ there is assumed to be random, independent of $\{Y_{ij}\}$. Evidently, $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are not independent in this case. Yet, Silverstein's result (1995) indicates that (2.5) still holds if $\{Y_{ij}\}$ are independent and independent of $\mathbf{T}_1$. This indicates that there may be the cases where the LSD of sample covariance matrix is also determined by (2.5) even when $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are not independent. A nature concern is whether the statistic L_n works in this case. We now consider the case when the random $\mathbf{T}_1$ is the inverse of another sample covariance matrix ($\mathbf{S}$ becomes the F matrix in this case). It is then proved in Theorem 3.1 of Zheng (2012) that L_n has a central limit theorem different from that for independent $\mathbf{x}_1, \cdots, \mathbf{x}_p$. The difference is caused by randomness of $\mathbf{T}_1$ and one may refer to (6.32) in Step 2 of Zheng's proof.

Although it is difficult to provide a central limi theorem for the statistic L_n for the general alternative hypothesis $\mathbb{H}_1$, we can still evaluate the power of L_n for a class of local alternatives. Specifically speaking, we consider a kind of local alternative with a sample covariance matrix in the form of $\mathbf{X}\mathbf{T}_2\mathbf{X}^T$, as in (2.6). Set

$$R_j^{(i)} = p\int x^i d\Big(F_{\mathbb{H}_1}^{\mathbf{S}^{(j)}}(x) - F_{\mathbb{H}_0}^{\mathbf{S}^{(j)}}(x)\Big),\ i=1,2;\ j=1,2; \tag{3.14}$$

where $F_{\mathbb{H}_0}^{\mathbf{S}^{(j)}}$ stands for the ESD of $\mathbf{S}^{(j)}$ under $\mathbb{H}_0$ while $F_{\mathbb{H}_1}^{\mathbf{S}^{(j)}}$ is the ESD of $\mathbf{S}^{(j)}$ under $\mathbb{H}_1$.

Theorem 4. *In addition to assumptions in Theorem 2 or Theorem 1, suppose that in probability*

$$\lim_{n\to\infty}\left|R_j^{(i)}\right| \to \infty, \quad \textit{for any } i, j. \tag{3.15}$$

Then

$$\lim_{n\to\infty} P(L_n > \gamma_{1-\alpha}|\mathbb{H}_1) = 1,$$

where $\gamma_{1-\alpha}$ is the critical value of χ^2 under $\mathbb{H}_0$ corresponding to the significance level α.

Remark 4. *Suppose that each column of $\mathbf{X}$ satisfies either Assumption 1 or Assumption 2 and all columns are independent. Condition (3.15) is equivalent to requiring*

$$tr\Big(\mathbf{X}^{(j)}\mathbf{T}^{(j)}(\mathbf{X}^{(j)})^T\Big)^i - tr\Big(\mathbf{X}^{(j)}(\mathbf{X}^{(j)})^T\Big)^i \to \infty, \quad \textit{for any } i, j \tag{3.16}$$

in probability, where $\mathbf{X}^{(j)}\mathbf{T}^{(j)}(\mathbf{X}^{(j)})^T$ denotes the sample covariance matrix of the jth group of the observations under the alternative $\mathbb{H}_1$ with $\mathbf{T}^{(j)}$ characterizing the dependence among observations, while $\mathbf{X}^{(j)}(\mathbf{X}^{(j)})^T$ stands for the sample covariance matrix of the jth group of the observations under the null hypothesis H_0.

If

$$\mathbf{T}^{(j)} = \mathbf{I} + \mathbf{e}\mathbf{e}^T,$$

where the elements of the vector $\mathbf{e}$ are all equal to one, then it is straightforward to verify that (3.16) is true. Moreover, most of the examples given in the subsequent section satisfy (3.16).

4 Simulation results

This section provides some simulated examples to show the finite sample performance of the proposed test statistic L_n. To show the efficiency of our test, some classical time series models, such as MA(1), AR(1) and ARMA(1,1) processes, are considered. As for the dependent structures, we consider some dependent structures described by MA(1) model, AR(1) model, ARMA(1,1) model and factor model. The factor model is commonly used to illustrate cross-sectional dependence in cross-sectional panel data analysis.

4.1 Empirical sizes and empirical powers

First we introduce the method of calculating empirical sizes and empirical powers. Since the asymptotic distribution of the proposed test statistic L_n is a classical distribution, i.e. χ^2 distribution of degree 2, the empirical sizes and powers are easy to calculate. Let $z_{1-\frac{1}{2}\alpha}$ be the $100(1-\frac{1}{2}\alpha)\%$ quantile of the asymptotic null distribution $\chi^2(2)$ of the test statistic L_n. With K replications of the data set simulated under the null hypothesis, we calculate the empirical size as

$$\hat{\alpha} = \frac{\{\sharp \ of \ L_n^H \geq z_{1-\frac{1}{2}\alpha}\}}{K}, \tag{4.1}$$

where L_n^H represents the value of the test statistic L_n based on the data simulated under the null hypothesis.

In our simulation, we choose $K = 1000$ as the number of repeated simulations. The significance level is $\alpha = 0.05$. Since the asymptotic null distribution of the test statistic is a classical distribution, the quantile $z_{1-\frac{1}{2}\alpha}$ is easy to know. Similarly, the empirical power is calculated as

$$\hat{\beta} = \frac{\{\sharp \ of \ L_n^A \geq z_{1-\frac{1}{2}\alpha}\}}{K}, \tag{4.2}$$

where L_n^A represents the value of the test statistic L_n based on the data simulated under the alternative hypothesis.

4.2 Testing independence

In order to derive independent stationary time series $\{\mathbf{x}_i = (X_{1i}, X_{2i}, \ldots, X_{pi})' : i = 1, \ldots, n\}$, we generate data from the following three data generating processes (DGPs):

$$DGP1: \ X_{ji} = Z_{ji} + \theta_1 Z_{j-1,i}, \ j = 1, 2, \ldots, p; \ i = 1, 2, \ldots, n; \tag{4.3}$$

$$DGP2: \ X_{ji} = \phi_1 X_{j-1,i} + Z_{ji}, \ j = 1, 2, \ldots, p; \ i = 1, 2, \ldots, n; \tag{4.4}$$

$$DGP3: \ X_{ji} - \phi_1 X_{j-1,i} = Z_{ji} + \theta_1 Z_{j-1,i}, \ j = 1, 2, \ldots, p; \ i = 1, 2, \ldots, n, \tag{4.5}$$

where $\{X_{0i}, Z_{ji} : j = 1, 2, \ldots, p; i = 1, 2, \ldots, n\} \sim i.i.d \ N(0, 1)$. For each DGP, we generate $p + 100$ observations and then discard the first 100 data in order to mitigate the impact of the initial values.

With these simulated data, we apply the proposed statistic L_n and calculate the empirical sizes. Table 1, Table 3 and Table 5 establish the empirical sizes for the three DGPs under different pairs of (p, n). The results show that our statistic L_n works well under the null hypothesis $\mathbb{H}_0$. Additionally, their empirical sizes from the bootstrap method proposed in Remark 3 are illustrated in Table 2, Table 4 and Table 6 respectively.

4.3 Testing dependence

4.3.1 Three types of correlated structures

In this section, we test four dependent structures with the proposed test and provide the powers under each case. As in the last part of this section, we first generate data $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ under DGP 1. To describe the cross-sectional dependence between $\mathbf{x}_{i_1}$ and $\mathbf{x}_{i_2}$, $\forall i_1 \neq i_2$, we generate new data $\widetilde{\mathbf{X}} = \mathbf{XT}$, where $\mathbf{T}$ is a $p \times p$ Hermitian matrix which is the square root of a covariance matrix. $\mathbf{T}$ is constructed by the following three methods.

1. MA(1) type covariance matrix $\mathbf{\Sigma}_{MA} = (\sigma_{kh}^{MA})_{k,h=1}^p$:

$$\sigma_{kh}^{(MA)} = \begin{cases} (1+\theta^2), & k = h; \\ \theta, & |k-h| = 1; \\ 0, & |k-h| > 1. \end{cases} \tag{4.6}$$

   Under this case, $\mathbf{T} = \mathbf{\Sigma}_{MA}^{1/2}$.

2. AR(1) type covariance matrix $\mathbf{\Sigma}_{AR} = (\sigma_{kh}^{(AR)})_{k,h=1}^p$:

$$\sigma_{kh}^{(AR)} = \frac{\phi^{|k-h|}}{1-\phi^2}. \tag{4.7}$$

   Under this case, $\mathbf{T} = \mathbf{\Sigma}_{AR}^{1/2}$.

3. ARMA(1,1) type covariance matrix $\mathbf{\Sigma}_{ARMA} = (\sigma_{kh}^{(ARMA)})_{k,h=1}^p$:

$$\sigma_{kh}^{(ARMA)} = \begin{cases} 1 + \frac{(\phi+\theta)^2}{1-\phi^2}, & k = h; \\ \phi + \theta + \frac{(\phi+\theta)^2\phi}{1-\phi^2}, & |k-h| = 1; \\ \phi^{|k-h|-1}(\phi + \theta + \frac{(\phi+\theta)^2\phi}{1-\phi^2}), & |k-h| \geq 2. \end{cases} \tag{4.8}$$

   Under this case, $\mathbf{T} = \mathbf{\Sigma}_{ARMA}^{1/2}$.

The powers under the three cases are illustrated in Table 7, Table 8 and Table 9. The true parameters are taken as $\phi = 0.8$ and $\theta = 0.2$. It can be seen that the powers are near 1 as n and p tend to infinity in the same order.

4.3.2 Factor model dependence

We consider a data generating process which comes from a dynamic factor model, which is always used to describe cross-sectional dependence.

$$X_{ji} = \boldsymbol{\lambda}^{'}\mathbf{f}_j + \varepsilon_{ji},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, p, \tag{4.9}$$

with

$$\mathbf{f}_j = \mathbf{z}_j + \theta\mathbf{z}_{j-1},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, p, \tag{4.10}$$

where $\boldsymbol{\lambda}$ is an $r \times 1$ deterministic vector whose elements are called factor loadings; $\mathbf{f}_j$ is an $r \times 1$ random vector generated from (4.10), whose elements are called factors and the cross-section dependence between $\mathbf{x}_{i_1}$ and $\mathbf{x}_{i_2}$ are caused by the common factors $\mathbf{f}_j$. $\{\mathbf{z}_j : j = 1, 2, \ldots, p\} \sim i.i.d\ N(\mathbf{0}_r, \mathbf{I}_r)$ where $\mathbf{0}_r$ is an $r \times 1$ vector with elements 0 and $\mathbf{I}_r$ is an $r \times r$ identity matrix. $\{\varepsilon_{ji} : j = 1, 2, \ldots, p; i = 1, 2, \ldots, n\} \sim i.i.d\ N(0, 1)$ are idiosyncratic errors.

First, we generate the factor loadings in the vector $\boldsymbol{\lambda}$ from $N(4, 1)$ before generating data from (4.9) and (4.10). After generating the data, we can apply the proposed test statistic L_n to the data and the empirical powers are shown in Table 10. From this table, we can see that the powers increase as the number of factors r increases. This is reasonable in the sense that more factors should bring in stronger dependence.

4.3.3 Common random dependence

We consider a special dependent structure which is caused by a common random part. The data generating process is as follows.

$$\mathbf{x}_i = \mathbf{A}\mathbf{y}_i, \quad i = 1, 2, \ldots, n, \tag{4.11}$$

where $\mathbf{A}$ is a $p \times p$ random matrix whose components are i.i.d standard normal random variables; and $\mathbf{y}_i$, $i = 1, 2, \ldots, n$ are independent $p \times 1$ random vectors, whose components are assumed to be i.i.d standard normal random variables.

Therefore the random vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are dependent due to the common random part $\mathbf{A}$. The empirical powers are listed in Table 11. From the table, we can see that the proposed statistic L_n is powerful to capture this kind of dependence.

4.3.4 ARCH type dependence

It is known that dependent relations may be linear dependence or nonlinear dependence. The examples above are all linear dependent structures. In this section, we will present a nonlinear dependent structure.

Let us consider an autoregressive conditional heteroskedasticity (ARCH) model of the form:

$$X_{ji} = Z_{ji}\sqrt{\alpha_0 + \alpha_1 X_{j,i-1}^2}, i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, p; \tag{4.12}$$

where $\{Z_{ji} : j = 1, 2, \ldots, p; i = 1, 2, \ldots, n\}$ are white noise error terms with zero mean and unit variance. Here we take $\alpha_0, \alpha_1 \in (0, 1)$ and $3\alpha_1^2 < 1$, since the fourth moment of the elements of X_{ji} exists.

From this model, the sequences $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ are dependent but uncorrelated. Moreover, this sequence is a multiple martingale difference sequence. The components of each vector $\mathbf{x}_i$ are independent here. This simplified assumption is imposed because the asymptotic theory is established for covariance time series under the assumption that the fourth moment equals 3 while the asymptotic theorem is provided for random vectors with i.i.d. components without this restriction.

Simulation results indicate that the proposed test statistic L_n can not detect this type of dependence between $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Nevertheless, if we replace the elements X_{jt} by X_{jt}^2, then our statistic L_n can capture the dependence of this type. This efficiency is due to the correlation between the high powers of $\{X_{jt} : t = 1, 2, \ldots, n\}$.

Table 12 lists the powers of the proposed statistics L_n testing model (4.12) in several cases, i.e. α_0 and α_1 take different values. From the table, we can find the phenomenon that as α_1 increases, the powers also increase. This is consistent with our intuition that larger α_1 brings about larger correlation between $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

5 An empirical application

We now apply the proposed method to the daily returns of the 96 stocks from S&P500, one of the most popular stock markets. The original data are the daily closed stock prices of the companies belonging to S&P500 from January 2011 to December 2011, with total 252 prices for each stock. The price for stock i at day j is denoted as S_{ji}. These data are derived from Wharton Research Data Services (WRDS). We use the logarithmic difference $X_{ji} = \ln(S_{ji}/S_{j-1,i})$. Then $N = 251$ logarithmic differences are available for each stock. Note that although we have $N = 251$ observations available for each stock here, we only use the first $p(p \le N)$ data to do the test. The value of p is comparable to n.

The interest here is to test whether the daily returns for the investigated n stocks are dependent. Here we study three groups of companies, i.e. $n = 60, 70, 90$ stocks respectively from S&P500. Since the distribution of $X_{j\tau}$ possesses high peak and heavy tails compared with the normal distribution, which is a typical property of the financial data (Rama (2001)), for simplicity we suppose that a transformation of the data follows a standard normal distribution,

$$\hat{X}_{ji} := \Big(\frac{X_{ji} - a_i}{b_i}\Big)^{\beta_j} \sim N(0, 1), \tag{5.1}$$

where a_i, b_i, β_i are unknown parameters. Figure 1 illustrates the smoothed empirical densities of the transformed data $\hat{X}_{ji}$ for all the selected 96 stocks under investigation. From these graphs, we can see that the model (5.1) is fitted well.

It is time to calculate L_n. We randomly choose n companies from the total available 96 companies and calculate the proposed statistic L_n. Repeat this experiment $k = 5$ times and obtain 5 values for L_n. They are listed in Table 13. From this table, we can see that more companies involved in the test lead to larger statistic values. For each case, all the five statistic values are outside the interval with critical values as the end points. We should reject the null hypothesis that the randomly chosen $n = 60, 70, 90$ stocks are independent at the significance level 5%. This coincides with the popular financial theory that states that cross-sectional dependence exists in modern stock markets.

6 Conclusion

This paper provides a novel approach for independence test among a large number random vectors including covariance stationary time series of length p by using the empirical spectral distribution of the sample covariance matrix of the grouped time series under investigation. This test can capture various kinds of dependent structures, e.g. MA(1) model, AR(1) model, ARCH(1) model and the dynamic factor model established in the simulation section. The conventional method(LRT proposed by Anderson (1984)) utilized the correlated relationship between random vectors with i.i.d components to capture their dependence, instead of covariance stationary time series. Hong (1996) proposed a test statistic based on correlation functions to investigate independence between two covariance stationary time series. On the one hand, this idea is only efficient for normal distributed data. It may be an inappropriate tool for non-Gaussian distributed data, such as martingale difference sequences (e.g.ARCH(1) model), nonlinear MA(1) model etc., which possess dependent but uncorrelated structures. On the other hand, his method is only applicable to independence test for finite number of covariance stationary time series. Then the proposed test is more advantageous in these two points. The simulation results and an empirical application to cross-sectional independence test for stock prices in S&P500 highlight this approach.

7 Appendix

In this appendix, we present some lemmas and technical facts used in the proofs of the main theorems.

7.1 Useful lemmas

Lemma 2 (Stein's equation)**.** *Let $\boldsymbol{\eta} = \{\eta_\ell\}_{\ell=1}^p$ be independent Gaussian random variables of zero mean, and $\Phi : \mathbb{R}^p \to \mathbb{C}$ be a differentiable function with polynomially bounded partial derivatives $\Phi'_\ell, \ell = 1, \dots, p$. Then we have*

$$E\{\eta_\ell \Phi(\boldsymbol{\eta})\} = E\{\eta_\ell^2\} E\{\Phi'_\ell(\boldsymbol{\eta})\},\ \ell = 1, \dots, p; \tag{7.1}$$

and

$$Var\{\Phi(\boldsymbol{\eta})\} \le \sum_{\ell=1}^{p} E\{\eta_\ell^2\} E\{|\Phi'_\ell(\boldsymbol{\eta})|^2\}. \tag{7.2}$$

Lemma 3 (Generalized Stein's equation of Lytova and Pastur (2009))**.** *Let η be a random variable such that $E\{|\eta|^{q+2}\} < \infty$ for a certain nonnegative integer q. Then for any function $\Phi : \mathbb{R} \to \mathbb{C}$ of the class C^{q+1} with bounded derivatives $\Phi^{(\ell)}, \ell = 1, \dots, q+1$, we have*

$$E\{\eta \Phi(\eta)\} = \sum_{\ell=0}^{q} \frac{\kappa_{\ell+1}}{\ell!} E\{\Phi^{(\ell)}(\eta)\} + \varepsilon_q, \tag{7.3}$$

where the remainder term ε_q admits the bound

$$|\varepsilon_q| \le C_q E(|\eta|^{q+2}) \sup_{t \in \mathbb{R}} |\Phi^{(q+1)}(t)|, \tag{7.4}$$

or

$$|\varepsilon_q| \leq C_q \int_0^1 E\big|\eta^{q+2}\Phi^{(q+1)}(\eta v)\big|(1-v)^q dv, \tag{7.5}$$

with $C_q \leq \frac{1+(3+2q)^{q+2}}{(q+1)!}$.

We would point out that (7.5) can be obtained from the proof of Lytova and Pastur (2009).

Our proof utilizes the generalized Fourier transform as follows:

Lemma 4 (Proposition 2.1 of Lytova and Pastur (2009)). *Let* $g : \mathbb{R}_+ \to \mathbb{C}$ *be locally Lipshitzian and such that for some* $\delta > 0$

$$\sup_{t\geq 0} e^{-\delta t}|g(t)| < \infty$$

and let $\widetilde{g} : \{z \in \mathbb{C} : Im(z) < -\delta\} \to \mathbb{C}$ *be its generalized Fourier transform*

$$\widetilde{g}(z) = i^{-1}\int_0^\infty e^{-izt}g(t)dt.$$

The inversion formula is given by

$$g(t) = \frac{i}{2\pi}\int_L e^{izt}\widetilde{g}(z)dz, t \geq 0,$$

where $L = (-\infty - i\varepsilon, \infty - i\varepsilon)$, $\varepsilon > \delta$, *and the principal value of the integral at infinity is used.*

Denote the correspondence between functions and their generalized Fourier transforms as $g \leftrightarrow \widetilde{g}$. Then we have

$$g^{'}(t) \leftrightarrow i\big(g(+0) + z\widetilde{g}(z)\big); \int_0^t g(\tau)d\tau \leftrightarrow (iz)^{-1}\widetilde{g}(z);$$
$$\int_0^t g_1(t-\tau)g_2(\tau)d\tau := (g_1 * g_2)(t) \leftrightarrow i\widetilde{g_1}(z)\widetilde{g}_2(z). \tag{7.6}$$

Furthermore, we introduce a simple fact about exponential matrices below.

Lemma 5 (Duhamel formula). *Let* $\mathbf{W}_1, \mathbf{W}_2$ *be* $n \times n$ *matrices and* $t \in \mathbb{R}$. *Then we have*

$$\mathbf{e}^{(\mathbf{W}_1+\mathbf{W}_2)t} = \mathbf{e}^{\mathbf{W}_1 t} + \int_0^t \mathbf{e}^{\mathbf{W}_1(t-s)}\mathbf{W}_2 e^{(\mathbf{W}_1+\mathbf{W}_2)s}ds. \tag{7.7}$$

Moreover, if $\big(W_{ij}(t)\big)_{1\leq i,j\leq n}$ *is a matrix-valued function of* $t \in \mathbb{R}$ *that is* C^∞ *in the sense that each matrix element* $W_{ij}(t)$ *is* C^∞. *Then*

$$\frac{d}{dt}\mathbf{e}^{\mathbf{W}(t)} = \int_0^1 \mathbf{e}^{s\mathbf{W}(t)}\mathbf{W}^{'}(t)\mathbf{e}^{(1-s)\mathbf{W}(t)}ds, \tag{7.8}$$

where $\mathbf{W}^{'}(t)$ *is an* $n \times n$ *matrix with elements being the derivatives of the corresponding elements of* $\mathbf{W}(t)$.

Proof of Theorem 1: Since

$$E\Big(\int \lambda dF^{\mathbf{S}}(\lambda)\Big) = E\Big(\frac{1}{p}tr(\frac{1}{n}\mathbf{X}\mathbf{X}')\Big) = \sum_{k=0}^{\infty} b_k^2,$$

the sequence $E\{F^{\mathbf{S}}(\lambda)\}$ is tight. By Theorem B.9 of Bai and Silverstein (2009), the proof of Theorem 1 is complete if we can verify the following two steps:

1. For any fixed $z \in \mathcal{C}^+$, $m_n(z) - Em_n(z) \to 0$, a.s. as $n \to \infty$, where $m_n(z) = \frac{1}{p}tr\mathbf{G}^{-1}(z)$ with $\mathbf{G}^{-1}(z) = (\mathbf{S} - z\mathbf{I}_p)^{-1}$ and $\mathbf{I}_p$ being a $p \times p$ identity matrix.

2. For any fixed $z \in \mathcal{C}^+$, $Em_n(z) \to m_\phi(z)$, as $n \to \infty$, where $m_\phi(z) = \int \frac{1}{\lambda - z} dF_{c,\phi}(\lambda)$.

The first step is omitted here, since it is similar to the proof on page 54 of Bai and Silverstein (2009).

We will finish the second step by comparing $Em_n(z)$ for the Gaussian case and nonGaussian case: as $n \to \infty$

$$Em_n(z) - E\hat{m}_n(z) \to 0, \tag{7.9}$$

$$E\hat{m}_n(z) \to m_\phi(z), \tag{7.10}$$

where $\hat{m}_n(z)$ is obtained from $m_n(z)$ with the elements $X_{jt} = \sum_{k=0}^{\infty} b_k \xi_{j-k,t}$ replaced by $\hat{X}_{jt} = \sum_{k=0}^{\infty} b_k \hat{\xi}_{j-k,t}$. Here $\{\hat{\xi}_{j-k,t}\}$ are i.i.d Gaussian random variables with mean zero and variance one and $\{\hat{\xi}_{j-k,t}\}$ are independent of $\{\xi_{j-k,t}\}$. (7.10) obviously holds by Yao (2012).

Let $Im(z) = v > 0$ and below we will frequently use the fact that $|\hat{m}_n(z)|$ and $|m_n(z)|$ are both bounded by $1/v$ without mention. We now consider (7.9) and start with the truncation of underlying random variables. Define

$$\mathbf{S}^\tau = \frac{1}{n}\mathbf{X}^\tau(\mathbf{X}^\tau)^T, \ \mathbf{X}^\tau = (X_{jt}^\tau)_{p\times n}, \ X_{jt}^\tau = \sum_{k=0}^{\infty} b_k \xi_{j-k,t}^\tau, \ \xi_{j-k,t}^\tau = \xi_{j-k,t} I(|\xi_{j-k,t}| \le \tau\sqrt{n}), \tag{7.11}$$

where $\tau = \tau_n$ is a positive sequence satisfying

$$\tau \to 0, \quad \frac{1}{\tau} E(|\xi_{11}|^2 I(|\xi_{11}| > \tau\sqrt{n})) \to 0. \tag{7.12}$$

We claim that for every $\tau > 0$,

$$\lim_{n\to\infty} \Big| Em_n(z) - Em_n^\tau(z) \Big| = 0, \tag{7.13}$$

where $m_n^\tau(z)=\frac{1}{p}tr\mathbf{G}_\tau^{-1}(z)$ with $\mathbf{G}_\tau^{-1}(z)=\frac{1}{p}tr(\mathbf{S}^\tau-z\mathbf{I}_p)^{-1}$. In fact, we have

$$\begin{aligned}
&\Big|Em_n(z)-Em_n^\tau(z)\Big| \\
\le\ &\Big|\frac{1}{p\sqrt{n}}\sum_{j,t=1}^{p,n}E\Big(\big(\mathbf{G}_\tau^{-1}(z)\mathbf{G}^{-1}(z)\frac{1}{\sqrt{n}}\mathbf{X}\big)_{jt}\big(X_{jt}-X_{jt}^\tau\big)\Big)\Big| \\
&+\Big|\frac{1}{p\sqrt{n}}\sum_{j,t=1}^{p,n}E\Big(\big(X_{jt}-X_{jt}^\tau\big)\big(\mathbf{G}^{-1}(z)\mathbf{G}_\tau^{-1}(z)\frac{1}{\sqrt{n}}\mathbf{X}\big)_{jt}\Big)\Big| \\
\le\ &\frac{Cnp}{p\sqrt{n}}\sum_{k=0}^{\infty}|b_k|E|\xi_{11}|I(|\xi_{11}|\ge\tau\sqrt{n})\le\frac{CE|\xi_{11}|^2I(|\xi_{11}|\ge\tau\sqrt{n})}{\tau}\sum_{k=0}^{\infty}|b_k|\to 0,
\end{aligned}$$

where the first inequality uses the resolvent identity

$$(\mathbf{A}-z\mathbf{I}_p)^{-1}-(\mathbf{B}-z\mathbf{I}_p)^{-1}=-(\mathbf{A}-z\mathbf{I}_p)^{-1}(\mathbf{A}-\mathbf{B})(\mathbf{B}-z\mathbf{I}_p)^{-1},$$

holding for any Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$ and the second inequality uses

$$\begin{aligned}
\Big|\Big(\mathbf{G}_\tau^{-1}(z)\mathbf{G}^{-1}(z)\frac{1}{\sqrt{n}}\mathbf{X}\Big)_{jt}\Big| &\le \frac{1}{v}||\mathbf{G}^{-1}(z)\frac{1}{\sqrt{n}}\mathbf{X}||=\frac{1}{v}||\mathbf{G}^{-1}(z)\frac{1}{n}\mathbf{X}\mathbf{X}^T\mathbf{G}^{-1}(z)||^{1/2} \\
&\le \frac{1}{v}||\mathbf{G}^{-1}(z)||^{1/2}+\frac{1}{v}|z|^{1/2}||\mathbf{G}^{-1}(z)||\le C. \qquad (7.14)
\end{aligned}$$

Here $\|\cdot\|$ denotes the spectral norm of a matrix. Also throughout the paper we use C to denote constants which may change from line to line.

In view of (7.13) it is sufficient to prove that$|Em_n^\tau(z)-E\hat{m}_n(z)|\to 0,\ as\ n\to\infty$. However for simplicity below we still use notation $m_n(z),\mathbf{X},X_{jt},\xi_{j-k,t}$ instead of using $m_n^\tau(z),\mathbf{X}^\tau,X_{jt}^\tau,\xi_{j-k,t}^\tau$ and prove (7.9). But one should keep in mind that $|\xi_{j-k,t}|\le\tau\sqrt{n}$.

We next prove (7.9) by an interpolation technique first introduced in Lytova and Pastur (2009). To this end define the interpolation matrix

$$\mathbf{S}(s)=\frac{1}{n}\mathbf{X}(s)\mathbf{X}^T(s),\mathbf{X}(s)=\Big(X_{\theta,t}(s)\Big)=s^{1/2}\mathbf{X}+(1-s)^{1/2}\hat{\mathbf{X}},\ s\in[0,1] \qquad (7.15)$$

and

$$\mathbf{G}^{-k}(s,z)=\Big(\mathbf{S}(s)-z\mathbf{I}_p\Big)^{-k},\ m_n(s,z)=\frac{1}{p}tr\mathbf{G}^{-1}(s,z),\quad k=1,2.$$

Write $\Phi_{jt}(s)=\Big(\mathbf{G}^{-2}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\Big)_{jt}$. We then have

$$Em_n(z)-E\hat{m}_n(z)=\int_0^1\frac{\partial}{\partial s}Em_n(s,z)ds=$$

$$-\frac{1}{p}\int_0^1\sum_{j,t=1}^{p,n}E\Big[s^{-1/2}\frac{1}{\sqrt{n}}X_{jt}\Phi_{jt}(s)\Big]ds+\frac{1}{p}\int_0^1\sum_{j,t=1}^{p,n}E\Big[(1-s)^{-1/2}\frac{1}{\sqrt{n}}\hat{X}_{jt}\Phi_{jt}(s)\Big]ds, \qquad (7.16)$$

where we have used the formula below

$$\frac{\partial \mathbf{G}^{-1}(s,z)}{\partial s} = -\mathbf{G}^{-1}(s,z)\frac{\partial \mathbf{S}(s)}{\partial s}\mathbf{G}^{-1}(s,z).$$

Consider the second term in (7.16) first. Since $\hat{X}_{jt} = \sum_{k=0}^{\infty} b_k \hat{\xi}_{j-k,t}$ we have

$$E\Big(\frac{1}{\sqrt{n}}\hat{X}_{jt}\Phi_{jt}(s)\Big) = \sum_{k=0}^{\infty} b_k E\Big(\frac{1}{\sqrt{n}}\hat{\xi}_{j-k,t}\Phi_{jt}(s)\Big). \tag{7.17}$$

Applying Lemma 2 to each summand in (7.17) we have

$$(1-s)^{-1/2}\sum_{k=0}^{\infty} b_k E(\frac{1}{\sqrt{n}}\hat{\xi}_{j-k,t}\Phi_{jt}(s)) = \frac{1}{n}\sum_{k=0}^{\infty} b_k \sum_{\theta=j-k}^{p} b_{\theta-j+k} E\Big(D_{\theta,t}(\Phi_{jt}(s))\Big), \tag{7.18}$$

where the partial derivative $D_{\theta,t} = \partial/\partial(\frac{1}{\sqrt{n}}X_{\theta t}(s))$ and we used the fact that

$$\frac{\partial \hat{X}_{\theta t}}{\partial \hat{\xi}_{j-k,t}} = b_{\theta-j+k}, \quad \frac{\partial X_{\theta t}(s)}{\partial \hat{X}_{\theta t}} = (1-s)^{1/2}.$$

Consider the first term in (7.16) now. As before, applying the fact that $X_{jt} = \sum_{k=0}^{\infty} b_k \xi_{j-k,t}$ and Lemma 3 to each summand of the first term in (7.16), we obtain

$$E\Big(s^{-1/2}\frac{1}{\sqrt{n}}X_{jt}\Phi_{jt}(s)\Big) = s^{-1/2}\sum_{k=0}^{\infty} b_k E\Big(\frac{1}{\sqrt{n}}\xi_{j-k,t}\Phi_{jt}(s)\Big) \tag{7.19}$$

$$= s^{-1/2}\frac{1}{\sqrt{n}}\sum_{k=0}^{\infty} b_k \kappa_{1,\tau} E\Phi_{jt}(s) + s^{-1/2}\frac{1}{n}\sum_{k=0}^{\infty} b_k \kappa_{2,\tau}\sum_{\zeta=j-k}^{p} b_{\zeta-j+k} E\Big(D_{\zeta,t}(\Phi_{jt}(s))\Big) + \varepsilon_1,$$

where $\kappa_{i,\tau}$ denotes the ith cumulant of the variable $\xi_{j-k,t}$ with $i = 1, 2$,

$$|\varepsilon_1| \le \frac{C_1 s^{-1/2}}{n^{3/2}}\sum_{k=0}^{\infty} |b_k| E\Big(|\xi_{j-k,t}|^3 \sup_{|\xi_{j-k,t}|\le \tau\sqrt{n}} |\widetilde{D}^2_{j-k,t}(\Phi_{jt}(s))|\Big),$$

with

$$\widetilde{D}^2_{j-k,t}(\Phi_{jt}(s)) = \widetilde{D}_{j-k,t}\Big(\sum_{\zeta=j-k}^{p} \frac{\partial \Phi_{jt}(s)}{\partial \frac{1}{\sqrt{n}}X_{\zeta,t}(s)}\frac{\partial \frac{1}{\sqrt{n}}X_{\zeta,t}(s)}{\partial \frac{1}{\sqrt{n}}X_{\zeta,t}}\frac{\partial \frac{1}{\sqrt{n}}X_{\zeta,t}}{\partial \frac{1}{\sqrt{n}}\xi_{j-k,t}}\Big)$$

$$= s\sum_{\zeta=j-k}^{p}\sum_{\gamma=j-k}^{p} b_{\zeta-j+k} b_{\gamma-j+k} D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big),$$

where $\widetilde{D}_{j-k,t} = \partial/\partial\frac{1}{\sqrt{n}}\xi_{j-k,t}$. Here we would point out that checking the argument of Lemma 3 in Lytova and Pastur (2009) shows that $\sup_{t\in R}$ in (7.5) can be replaced by $\sup_{|\xi_{j-k,t}|\le \tau\sqrt{n}}$ in the remainder ε_1 due to the truncation step.

We conclude from (7.16)-(7.19) that

$$Em_n(z) - E\hat{m}_n(z) = -\int_0^1 \Big[\frac{s^{-1/2}}{pn^{1/2}}\sum_{k=0}^{\infty} b_k \sum_{j,t=1}^{p,n} \kappa_{1,\tau} E\Phi_{jt}(s) + \frac{1}{p}\sum_{j,t=1}^{p,n} \varepsilon_1$$

$$+\frac{s^{-1/2}}{np}\sum_{k=0}^{\infty}b_k\sum_{j,t=1}^{p,n}(\kappa_{2,\tau}-1)\sum_{\zeta=j-k}^{p}b_{\zeta-j+k}E\Big(D_{\zeta,t}(\Phi_{jt}(s))\Big)\Big]ds. \tag{7.20}$$

The next aim is to prove that each of the three integrands goes to zero as n tends to infinity. To this end, first let $\mu_{\ell,\tau}(\mu_\ell)$ and $\kappa_{\ell,\tau}(\kappa_\ell)$ be the ℓth moment and cumulant of the truncated ξ_{jt} and the untruncated (ξ_{jt}) respectively. Then

$$|\mu_{\ell,\tau}-\mu_\ell|\le CE\Big(|\xi_{11}|^\ell I(|\xi_{11}|>\tau\sqrt{n})\Big).$$

As a result we have

$$|\kappa_{\ell,\tau}-\kappa_\ell|\le CE\Big(|\xi_{11}|^\ell I(|\xi_{11}|>\tau\sqrt{n})\Big)\le\frac{C}{(\tau\sqrt{n})^{2-\ell}}E(|\xi_{11}|^2I(|\xi_{11}|>\tau\sqrt{n})). \tag{7.21}$$

This result uses the fact that cumulants can be expressed by moments as follows

$$\kappa_j=\sum_\lambda c_\lambda\mu_\lambda,$$

where the sum is over all additive partitions λ of the set $\{1,\dots,j\}$, $\{c_\ell:\ell\in\lambda\}$ are known coefficients and $\mu_\lambda=\prod_{\ell\in\lambda}\mu_\ell$.

Second we provide the upper bound of $\Phi_{jt}(s)$, $D_{\gamma,t}(\Phi_{jt}(s))$ and $D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big)$. For simplicity, we introduce more new notation.

$$\begin{aligned}
&\mathbf{I}(\zeta,\gamma)=\mathbf{e}_\gamma\mathbf{e}_\zeta^T+\mathbf{e}_\zeta\mathbf{e}_\gamma^T,\quad \mathbf{W}(\gamma,t)=\mathbf{e}_\gamma\mathbf{e}_t^T\frac{1}{\sqrt{n}}\mathbf{X}^T(s)+\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t\mathbf{e}_\gamma^T,\\
&\mathbf{J}_1(\zeta)=\mathbf{G}^{-1}(s,z)\mathbf{W}(\zeta,t)\mathbf{G}^{-2}(s,z),\quad \mathbf{J}_2(\gamma,\zeta)=\mathbf{G}^{-1}(s,z)\mathbf{I}(\gamma,\zeta)\mathbf{G}^{-2}(s,z)\\
&\mathbf{J}_3(\gamma,\zeta)=\mathbf{G}^{-1}(s,z)\mathbf{W}(\gamma,t)\mathbf{G}^{-2}(s,z)\mathbf{W}(\zeta,t)\mathbf{G}^{-1}(s,z),\\
&\mathbf{J}_4(\zeta,\gamma)=\mathbf{G}^{-1}(s,z)\mathbf{W}(\zeta,t)\mathbf{G}^{-1}(s,z)\mathbf{W}(\gamma,t)\mathbf{G}^{-2}(s,z),
\end{aligned}$$

where $\mathbf{e}_\gamma$ and $\mathbf{e}_j$ are $p\times1$ unit vectors with the γ-th and j-th elements being 1 respectively and others being zeros; and $\mathbf{e}_t$ is $n\times1$ a unit vector with the t-th element being 1 and others being zeros. With these notation by a simple but tedious calculation we obtain

$$D_{\gamma,t}(\Phi_{jt}(s))\;=\;-\mathbf{e}_j^T\mathbf{G}^{-2}(s,z)\mathbf{e}_\gamma+\mathbf{e}_j^T\mathbf{J}_1(\gamma)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t+\mathbf{e}_j^T\mathbf{J}_1^T(\gamma)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t$$

and

$$\begin{aligned}
&D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big)=\mathbf{e}_j^T\mathbf{J}_1(\zeta)\mathbf{e}_\gamma+\mathbf{e}_j^T\mathbf{J}_1^T(\zeta)\mathbf{e}_\gamma-\mathbf{e}_j^T\mathbf{J}_4(\zeta,\gamma)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t\\
&-\mathbf{e}_j^T\mathbf{J}_4(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t-\mathbf{e}_j^T\mathbf{J}_3(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t+\mathbf{e}_j^T\mathbf{J}_1(\gamma)\mathbf{e}_\zeta-\mathbf{e}_j^T\mathbf{J}_1^T(\gamma)\mathbf{e}_\zeta\\
&-\mathbf{e}_j^T\mathbf{J}_3(\zeta,\gamma)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t-\mathbf{e}_j^T\mathbf{J}_4^T(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t-\mathbf{e}_j^T\mathbf{J}_4^T(\zeta,\gamma)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t\\
&+\mathbf{e}_j^T\mathbf{J}_2(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t+\mathbf{e}_j^T\mathbf{J}_2^T(\gamma,\zeta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\mathbf{e}_t.
\end{aligned}$$

From the expansions of $\Phi_{jt}(s), D_{\gamma,t}(\Phi_{jt}(s))$ and $D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big)$ we see that all the terms in such expansions include only three factors below:

$$D_1 = \Big(\frac{1}{\sqrt{n}}\mathbf{X}^T(s)\mathbf{G}^{-\ell}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\Big)_{tt},\ D_2 = \Big(\mathbf{G}^{-\ell}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\Big)_{kt},$$
$$D_3 = \mathbf{G}^{-\ell}(s,z)_{kk'},\quad \ell = 1,2, k, k' = j, \zeta,\ or\ \gamma.$$

These three factors turn out to be bounded, as seen below.

Obviously $|D_3| \le v^{-\ell}$. Similar to (7.14) using

$$\mathbf{G}^{-1}(z)\frac{1}{n}\mathbf{X}(s)\mathbf{X}^T(s) = I + z\mathbf{G}^{-1}(s,z). \tag{7.22}$$

one may verify that

$$|D_2| \le \frac{1}{v^{\ell-1}}\|\mathbf{G}^{-1}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\| \le C,\ j = 1,2$$

and

$$|D_1| \le \|\frac{1}{\sqrt{n}}\mathbf{X}^T(s)\mathbf{G}^{-\ell}(s,z)\frac{1}{\sqrt{n}}\mathbf{X}(s)\| = \| \mathbf{G}^{-\ell}(s,z)\frac{1}{n}\mathbf{X}(s)\mathbf{X}^T(s) \| \le C.$$

Therefore $\Phi_{jt}(s)$ and the two derivatives $D_{\gamma,t}(\Phi_{jt}(s))$, $D_{\zeta,t}\Big(D_{\gamma,t}(\Phi_{jt}(s))\Big)$ are bounded. This, together with (7.21) and (7.12), yields

$$\Big|\frac{s^{-1/2}}{pn^{1/2}}\sum_{k=0}^{\infty} b_k \sum_{j,t=1}^{p,n} \kappa_{1,\tau} E\Phi_{jt}\Big| \le \frac{C}{\tau}E(|\xi_{11}|^2 I(|\xi_{11}| > \tau\sqrt{n})) \to 0$$

and

$$\Big|\frac{1}{np}\sum_{k=0}^{\infty} b_k \sum_{j,t=1}^{p,n}(\kappa_{2,\tau}-1)\sum_{\zeta=j-k}^{p} b_{\zeta-j+k}E\Big(D_{\zeta,t}(s)\Phi_{jt}(s)\Big)\Big| \le CE(|\xi_{11}|^2 I(|\xi_{11}| > \tau\sqrt{n})) \to 0.$$

Moreover since $E|\xi_{jt}|^3 \le \tau\sqrt{n}$ and (7.12) we have

$$\Big|\frac{1}{p}\sum_{j,t=1}^{p,n}\varepsilon_1\Big| \le C\tau \to 0,\ as\ n \to \infty.$$

These, together with (7.20), yield (7.9). The proof of this theorem is complete. □

Proof of Theorem 2: The strategy of the proof is the same as that in Lytova and Pastur (2009). That is, we first establish CLT for the case when $\{\xi_{j-k,t}\}$ are i.i.d $N(0,1)$ and then generalize it to the general distributions.

When $\{\xi_{j-k,t}\}$ are i.i.d $N(0,1)$, as stated in Section 2, under $\mathbf{H}_0$, the matrix $\mathbf{S}$ can be written in the form that $\mathbf{S} = \frac{1}{n}\mathbf{T}_1^{1/2}\mathbf{X}\mathbf{X}^T\mathbf{T}_1^{1/2}$ so that Theorem 9.10 of Bai and Silverstein (2009) is applicable. The asymptotic variance of Theorem 2 is the same as that in Bai and Silverstein (2009) while the asymptotic mean is obtained from that in Bai and Silverstein (2009) and the facts that (See Yao (2012) and Gray (2009))

$$\lim_{p\to\infty}\frac{1}{p}\sum_{k=1}^{p}f(\sigma_k)=\int_0^{\infty}f(x)dH(x)=\frac{1}{2\pi}\int_0^{2\pi}f(\phi(\lambda))d\lambda.$$

However to apply Bai and Silverstein (2009), we have to make sure that the spectral norm of the population covariance matrix $\mathbf{T}_1$ of each time series is bounded. We claim that this is ensured by the condition $\sum|b_j|<\infty$. In fact, let $\sigma_k=Cov(X_{jt},X_{j+k,t})$. By the expression (2.2) of the time series and a change of variables we have

$$\begin{aligned}\sum_{k=0}^{\infty}|\sigma_k| &= \sum_{k=0}^{\infty}|Cov(\sum_{k_1=0}^{\infty}b_{k_1}\xi_{j-k_1,t},\sum_{k_2=0}^{\infty}b_{k_2}\xi_{j+k-k_2,t})| \\ &= \sum_{k=0}^{\infty}|\sum_{k_1=0}^{\infty}b_{k_1}b_{k_1+k}|<(\sum_{k=0}^{\infty}|b_k|)^2<\infty. \end{aligned}\tag{7.23}$$

By Lemma 4.1 of Gray (2009) and (7.23) we conclude that

$$||\mathbf{T}_1||\le 4\sum_{k=0}^{\infty}|\sigma_k|<\infty.\tag{7.24}$$

We next adopt an interpolation trick and compare the CLT of the general case with that of the Gaussian case. Recall the definition of $G_n(\lambda)$ in (2.8). Let

$$\mathcal{N}_n^{\circ}[f]=\int f(\lambda)dG_n(\lambda),\quad \mathcal{N}_n[f]=\int f(\lambda)dpF^{\mathbf{S}}(\lambda).$$

Define $\widehat{\mathcal{N}}_n^{\circ}[f]$ and $\widehat{\mathcal{N}}_n[f]$ to be obtained from $\mathcal{N}_n^{\circ}[f]$ and $\mathcal{N}_n[f]$ respectively, with the entries $X_{jt}=\sum_{k=0}^{\infty}b_k\xi_{j-k,t}$ replaced by $\hat{X}_{jt}=\sum_{k=0}^{\infty}b_k\hat{\xi}_{j-k,t}$ where $\{\hat{\xi}_{j-k,t}\}$ are i.i.d. N(0,1) and independent of $\{\xi_{j-k,t}\}$. By the continuous theorem of characteristic functions, it suffices to show that

$$R_n(x):=E\Big(e^{ix\mathcal{N}_n^{\circ}[f]}\Big)-E\Big(e^{ix\widehat{\mathcal{N}}_n^{\circ}[f]}\Big)\to 0,\ as\ n\to\infty.\tag{7.25}$$

Since the integrand function f admits the Fourier transform

$$\hat{f}(\theta)=\frac{1}{2\pi}\int e^{-i\theta\lambda}f(\lambda)d\lambda,$$

the Fourier inversion formula is

$$f(\lambda)=\int e^{i\theta\lambda}\hat{f}(\theta)d\theta.\tag{7.26}$$

Then the statistic $\mathcal{N}_n[f]$ can be written as

$$\mathcal{N}_n[f]=\int\hat{f}(\theta)u_n(\theta)d\theta,$$

where

$$u_n(\theta) = Tr\mathbf{U}(\theta),\ \mathbf{U}(\theta) = \mathbf{e}^{i\theta\mathbf{S}}. \tag{7.27}$$

By (7.26) we obtain

$$f^{'}(\mathbf{S}) = i\int \hat{f}(\theta)\theta\mathbf{U}(\theta)d\theta. \tag{7.28}$$

We still use the same truncation as that in (7.11) (and use the same notation) but this time τ satisfies (see formula (9.7.7) of Bai and Silverstein (2009))

$$\tau \to 0, \quad \tau^{-4}E|\xi_{j-k,t}|^4 I(\xi_{j-k,t}| > \tau\sqrt{n}) \to 0. \tag{7.29}$$

Note that

$$P\{\mathbf{X} \neq \mathbf{X}^{\tau}\} \leq \sum_{j,t=1}^{p,n} P\{X_{jt} \neq X_{jt}^{\tau}\} \leq \frac{1}{\tau^4 n^2}\sum_{j,t=1}^{p,n}\sum_{k=0}^{\infty} b_k E|\xi_{j-k,t}|^4 I(\xi_{j-k,t}| > \tau\sqrt{n}) \to 0.$$

In view of this it is enough to prove that

$$E\Big(e^{ix\mathcal{N}_{n\tau}^{\circ}[f]}\Big) - E\Big(e^{ix\widehat{\mathcal{N}}_{n}^{\circ}[f]}\Big) \to 0,\ as\ n \to \infty, \tag{7.30}$$

where $\mathcal{N}_{n\tau}^{\circ}[f]$ is obtained from $\mathcal{N}_{n}^{\circ}[f]$ with $\mathbf{X}$ replaced by $\mathbf{X}^{\tau}$.

As in the proof of Theorem 1 we still use notation $\xi_{j-k,t}, \mathbf{X}, \mathcal{N}_n[f]$ rather than $\xi_{j-k,t}^{\tau}, \mathbf{X}^{\tau}, \mathcal{N}_{n\tau}^{\circ}[f]$ and below prove (7.25). Recall the interpolation matrix defined in (7.15) and furthermore define

$$e_n(s,x) = \exp\Big(ixTrf\big(\mathbf{S}(s)\big)\Big), \quad \mathbf{U}(s,\theta) = (U_{jk}) = \mathbf{e}^{i\theta\mathbf{S}(s)}.$$

By (7.28) we have

$$\begin{aligned} R_n(x) = a_n\int_0^1 \frac{\partial}{\partial s}E\Big(e_n(s,x)\Big)ds & \\ = \ ixa_n\int_0^1 E\Big[e_n(s,x)Tr\Big(f^{'}(\mathbf{S}(s))\big(s^{-1/2}\frac{1}{\sqrt{n}}\mathbf{X} - (1-s)^{-1/2}\frac{1}{\sqrt{n}}\hat{\mathbf{X}}\big)\frac{1}{\sqrt{n}}\mathbf{X}^{\tau'}(s)\Big)\Big]ds & \\ = \ -xa_n\int_0^1 ds\int \theta\hat{f}(\theta)(D_n - B_n)d\theta, & \qquad (7.31) \end{aligned}$$

where $a_n = \exp(-ix\int fdpF_{c_n,\phi_n})$ and

$$D_n = \frac{1}{\sqrt{ns}}\sum_{j,t=1}^{p,n} E\Big(X_{jt}\Psi_{jt}(s)\Big),\ B_n = \frac{1}{\sqrt{n(1-s)}}\sum_{j,t=1}^{p,n} E\Big(\hat{X}_{jt}\Psi_{jt}(s)\Big),$$

with

$$\Psi_{jt}(s) = e_n(s,x)\Big(\mathbf{U}(s,\theta)\frac{1}{\sqrt{n}}\mathbf{X}(s)\Big)_{jt}.$$

By Lemma 2 and a calculation similar to (7.17), (7.18) and (7.19) we obtain

$$B_n = \frac{1}{n}\sum_{j,t=1}^{p,n}\sum_{k=0}^{\infty} b_k \sum_{k^{(1)}=j-k}^{p} b_{k^{(1)}-j+k} E\Big(D_{k^{(1)}t}(\Psi_{jt}(s))\Big), \tag{7.32}$$

where $D_{k^{(1)}t} = \partial/\partial \frac{1}{\sqrt{n}} X_{k^{(1)}t}$.

Also, by Lemma 3 with $q = 3$ we have

$$D_n = \sum_{\ell=0}^{3} T_{\ell\tau} + \varepsilon_3, \tag{7.33}$$

where

$$T_{0\tau} = \frac{s^{-1/2}}{\sqrt{n}} \sum_{j,t=1}^{p,n} \kappa_{1,\tau} \sum_{k=0}^{\infty} E\Psi_{jt}(s),$$

$$\begin{aligned} T_{\ell\tau} \;\; = \;\; & \frac{s^{(\ell-1)/2}}{\ell! n^{(\ell+1)/2}} \sum_{j,t=1}^{p,n} \kappa_{\ell+1,\tau} \sum_{k=0}^{\infty} b_k \sum_{k^{(\ell)},k^{(\ell-1)},\ldots,k^{(1)}}^{p} b_{k^{(\ell)}-j+k} b_{k^{(\ell-1)}-j+k} \cdots b_{k^{(1)}-j+k} \\ & \cdot E\Big(D_{k^{(\ell)}t} D_{k^{(\ell-1)}t} \cdots D_{k^{(1)}t} \Psi_{jt}(s)\Big), \;\; \ell = 1,2,3; \end{aligned}$$

and

$$\begin{aligned} |\varepsilon_3| \;\; \le \;\; & \frac{Cs^2}{n^{5/2}} \sum_{j,t=1}^{p,n}\sum_{k=0}^{\infty} |b_k| \sum_{k^{(4)},\ldots,k^{(1)}=j-k}^{p} |b_{k^{(4)}-j+k}| \cdots |b_{k^{(1)}-j+k}| \\ & \cdot \int_0^1 E\Big[|\xi_{j-k,t}|^5 D_{k^{(4)}t} \cdots D_{k^{(1)}t} \Psi_{jt}(s)\big|_{\xi_{j-k,t}=v\xi_{j-k,t}}\Big](1-v)^3 dv, \end{aligned} \tag{7.34}$$

where $\Psi_{jt}(s)\big|_{\xi_{j-k,t}=v\xi_{j-k,t}}$ means that $\xi_{j-k,t}$ involved in $\Psi_{jt}(s)$ is replaced by $v\xi_{j-k,t}$ and $\kappa_{\ell,\tau}$ is the ℓth cumulant of $\xi_{j-k,t}$.

Next, we provide the upper bounds of derivatives:

$$D_{k^{(\ell)}t} D_{k^{(\ell-1)}t} \cdots D_{k^{(1)}t} \Psi_{jt}(s), \;\; \ell = 0,1,2,3,4.$$

Let $\mathbf{Y}(s) = (Y_{rt}(s)) = \frac{1}{\sqrt{n}}\mathbf{X}(s)$. Applying the Duhamel formula of Lemma 4 to the entries, $U_{jk^{(\ell)}}$, of $\mathbf{U}(s,\theta)$ we have

$$D_{\beta\alpha}(U_{jk^{(\ell)}}) = i\Big[\Big((\mathbf{UY}(s))_{j\alpha} * U_{\beta,k^{(\ell)}}\Big)(\theta) + \Big((\mathbf{UY}(s))_{k^{(\ell)}\alpha} * U_{j\beta}\Big)(\theta)\Big], \tag{7.35}$$

where the convolution $*$ is defined in (7.6). Here and below we use $\mathbf{U}$ to denote $\mathbf{U}(s,\theta)$ when there is no confusion. In view of (7.35) and the fact that $\mathbf{I}_p = \sum_{r=1}^{p} \mathbf{e}_r\mathbf{e}_r'$ we have

$$\begin{aligned} & D_{k^{(\ell)}t}(\mathbf{UY}(s))_{jt} = D_{k^{(\ell)}t}\Big(\sum_{r=1}^{p} Y_{rt}(s) U_{rj}\Big) \\ = \;\; & U_{k^{(\ell)}j} + i\Big[\Big((\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt} * U_{k^{(\ell)}j}\Big)(\theta) + \Big((\mathbf{UY}(s))_{jt} * (\mathbf{UY}(s))_{k^{(\ell)}t}\Big)(\theta)\Big], \end{aligned} \tag{7.36}$$

$$D_{k^{(d)}t}(\mathbf{Y}^T(s)\mathbf{U}\mathbf{Y}(s))_{tt} = D_{k^{(d)}t}\Big(\sum_{r=1}^{p}(\mathbf{U}\mathbf{Y}(s))_{rt}\mathbf{Y}_{rt}(s)\Big)$$

$$= \quad 2(\mathbf{U}\mathbf{Y}(s))_{k^{(d)}t} + 2i\Big((\mathbf{Y}^{T'}(s)\mathbf{U}\mathbf{Y}(s))_{tt} * (\mathbf{U}\mathbf{Y}(s))_{k^{(d)}t}\Big)(\theta), \tag{7.37}$$

and by (7.28)

$$D_{k^{(\ell)}t}(e_n(s,x)) = -2xe_n(s,x)\int \theta\hat{f}(\theta)(\mathbf{U}\mathbf{Y}(s))_{k^{(\ell)}t}d\theta, \tag{7.38}$$

where $\ell, d = 1,2,3,4$.

Since $\sum_{t=1}^{n}|U_{\alpha t}|^2 = 1$ and $||\mathbf{U}|| = 1$, from Hölder's inequality, we obtain

$$|(\mathbf{U}\mathbf{Y}(s))_{jt}| \le \Big(\sum_{r=1}^{p}(Y_{rt}(s))^2\Big)^{1/2}, \ |(\mathbf{Y}^T(s)\mathbf{U}\mathbf{Y}(s))_{tt}| \le \sum_{r=1}^{p}(Y_{rt}(s))^2. \tag{7.39}$$

Recalling the definition of $\Psi_{jt}(s)$ and repeatedly using (7.35)-(7.39) one can verify that

$$\Big|D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)\Big| \le C + C\Big(\sum_{r=1}^{p}(Y_{rt}(s))^2\Big)^{(\ell+1)/2}, \ \ell = 0,1,2,3,4. \tag{7.40}$$

For example see (7.49) below for the expansion of $D_{k^{(1)}t}\Psi_{jt}(s)$. Moreover it is straightforward to check that $\ell = 0,1,2,3$, $E\big(\sum_{r=1}^{p}(Y_{rt}(s))^2\big)^{(\ell+1)/2}$ is bounded by the fact that $n^2E|Y_{rt}(s)|^4 = E|X_{rt}(s)|^4 < \infty$. We then conclude that

$$E\Big|D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)\Big| \le C_\ell, \ \ell = 0,1,2,3. \tag{7.41}$$

However, to prove $\varepsilon_3 \to 0$, (7.40) for the case $\ell = 4$ is not enough for our purpose since

$$E|X_{rt}(s)|^5 \le C\tau\sqrt{n}, \tag{7.42}$$

not bounded. To offset this $\sqrt{n}$, one key observation is that from (7.35)-(7.38) we see that each term in the expansion of $D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)$ is a product or a convolution of some of the following factors

$$(\mathbf{U}\mathbf{Y}(s))_{h_1t}, \ (\mathbf{U})_{h_2h_3}, \ (\mathbf{Y}^T(s)\mathbf{U}\mathbf{Y}(s))_{tt}, \ e_n(s,x),$$

where h_i can be j or any $k^{(\ell)}$, $\ell = 1,\cdots,4$. Let m_1 and m_2 be the total number of factors of types of $(\mathbf{U}\mathbf{Y}(s))_{h_1t}$ and $(\mathbf{Y}^T(s)\mathbf{U}\mathbf{Y}(s))_{tt}$ appearing in each term of the expansion, respectively. Then from (7.35)-(7.38) and (7.49) below we see that $(m_1+2m_2) \le 5$ (this explains (7.40) to some extent). Consider the case when $(m_1+2m_2) = 5$ first. In this case from (7.35)-(7.38) and (7.49) below we see that at least one $(\mathbf{U}\mathbf{Y}(s))_{h_1t}$ must be contained in the expansion. We below show how to handle such terms by demonstrating one example and all other cases can be similarly proved. Consider the term

$$(\mathbf{U}\mathbf{Y}(s))_{jt}(\mathbf{U})_{k^{(2)}k^{(3)}}(\mathbf{U})_{k^{(4)}k^{(1)}}(\mathbf{Y}^T(s)\mathbf{U}\mathbf{Y}(s))_{tt}^2 \tag{7.43}$$

($m_1 = 1$ and $m_2 = 2$ in this case). Then for (7.34), it can be estimated as follows

$$
\begin{aligned}
&\frac{1}{n^{5/2}}\sum_{j,t=1}^{p,n}\int_0^1 E\Big[|\xi_{j-k,t}|^5\Big((\mathbf{UY}(s))_{jt}(\mathbf{U})_{k^{(2)}k^{(3)}}(\mathbf{U})_{k^{(4)}k^{(1)}}(\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt}^2\Big)\big|_{\xi_{j-k,t}=v\xi_{j-k,t}}\Big]\\
&\times\ (1-v)^3dv\\
= &\frac{1}{n^{5/2}}\sum_{j,t=1}^{p,n}\int_0^1 E\Big[|\eta|^5\Big((\mathbf{UY}(s))_{jt}(\mathbf{U})_{k^{(2)}k^{(3)}}(\mathbf{U})_{k^{(4)}k^{(1)}}(\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt}^2\Big)\big|_{\xi_{j-k,t}=v\eta}\Big](1-v)^3dv\\
\le &\frac{1}{n^{5/2}}\sum_{t=1}^{n}\int_0^1 E\Big[|\eta|^5\Big(\sum_{j=1}^{p}|(\mathbf{UY}(s))_{jt}||(\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt}|\Big)\big|_{\xi_{j-k,t}=v\eta}\Big](1-v)^3dv\\
\le &\frac{1}{n^{5/2}}\sum_{t=1}^{n}\int_0^1 E\Big[|\eta|^5\sqrt{p}\Big((\sum_{r=1}^{p}Y_{rt}^2)^{5/2}\Big)\big|_{\xi_{j-k,t}=v\eta}\Big](1-v)^3dv, \qquad (7.44)
\end{aligned}
$$

where η has the same distribution as $\{\xi_{r-k,t}\}$ and is independent of them, and satisfies $|\eta| \le \tau\sqrt{n}$; the first inequality uses the fact that $|(\mathbf{U})_{h_1h_2}| \le 1$; and the second inequality uses the second inequality of (7.39) and the following estimation

$$
\begin{aligned}
\sum_{j=1}^{p}|(\mathbf{UY}(s))_{jt}| &\le \sqrt{p}\Big(\sum_{j=1}^{p}|(\mathbf{UY}(s))_{jt}|^2\Big)^{1/2} = \sqrt{p}\Big(\sum_{j=1}^{p}\mathbf{e}_t^T\mathbf{Y}^T(s)\mathbf{U}^T\mathbf{e}_j\mathbf{e}_j^T\bar{\mathbf{U}}\mathbf{Y}(s)\mathbf{e}_t\Big)^{1/2}\\
&= \sqrt{p}\Big(\mathbf{e}_t^T\mathbf{Y}^T(s)\mathbf{Y}(s)\mathbf{e}_t\Big)^{1/2} = \sqrt{p}\Big(\sum_{r=1}^{p}Y_{rt}^2(s)\Big)^{1/2}, \qquad (7.45)
\end{aligned}
$$

where the second equality uses the fact that $\mathbf{U}$ is a symmetric unitary matrix. Moreover, since for any $h = 1, 2, \ldots, p$, the coefficient of $v\eta$ in the expansion of $Y_{rt}\big|_{\xi_{j-k,t}=v\eta}$ is b_{r-j+k} when $\xi_{j-k,t}$ is replaced by $v\eta$, we have

$$
\begin{aligned}
\Big(\sum_{r=1}^{p}Y_{rt}^2(s)\big|_{\xi_{j-k,t}=v\eta}\Big)^{m/2} &\le \frac{C}{n^{m/2}}\Big((\sum_{r=1}^{p}|b_{r-h+k}|)^m(\tau\sqrt{n})^m + (\sum_{r=1}^{p}\widetilde{X}_{rt}^2(s))^{m/2}\Big)\\
&\le C + \frac{C}{n}\sum_{r=1}^{p}\widetilde{X}_{rt}^m(s),\ \ 2 \le m \le 5, \qquad (7.46)
\end{aligned}
$$

where $\widetilde{X}_{rt}(s)$ is $X_{rt}(s) = s^{1/2}\sum_{\ell=0}^{\infty}b_\ell\xi_{r-\ell,t} + (1-s)^{1/2}\hat{X}_{rt}$ without the factor $\xi_{j-k,t} = v\eta$; and the last inequality utilizes the condition that $\sum_{\ell=0}^{\infty}|b_\ell| < \infty$. Note that $\widetilde{X}_{rt}$ is independent of η. This, together with (7.46) and the fact$E(|\widetilde{X}_{rt}|^5(s)) \le C\tau\sqrt{n}$, implies that

$$(7.44) \le C\tau \to 0.$$

If $(\mathbf{UY}(s))_{jt}$ in (7.43) is replaced by any $(\mathbf{UY}(s))_{k^{(i)}t}, i = 1, 2, 3, 4$, then an estimate similar to (7.44) also holds by exchanging the order of summation as follows

$$\sum_{j=1}^{p}\sum_{k=0}^{\infty}\sum_{k^{(i)}=j-k}^{p} b_{k^{(i)}-j+k}(\mathbf{UY}(s))_{k^{(i)}t} = \sum_{k=0}^{\infty}\sum_{k^{(i)}=1-k}^{p}(\mathbf{UY}(s))_{k^{(i)}t}\sum_{j=1}^{k^{(i)}+k} b_{k^{(i)}-j+k}. \qquad (7.47)$$

Next consider the case when $(m_1+m_2)\le 4$. By (7.40) and (7.46), one may verify that

$$\frac{1}{n^{5/2}}\sum_{j,t=1}^{p,n}\int_0^1 E\Big[|\xi_{j-k,t}|^5\Big((\mathbf{UY}(s))_{h_1t}^{m_1}(\mathbf{U})_{k^{(2)}k^{(3)}}^{m_3}(\mathbf{U})_{k^{(4)}k^{(1)}}^{m_4}(\mathbf{Y}^T(s)\mathbf{UY}(s))_{tt}^{m_2}\Big)\big|_{\xi_{j-k,t}=v\xi_{j-k,t}}\Big](1-v)^3dv$$

$$\le C\tau\to 0,$$

where $m_i\ge 0, i=3,4$. Summarizing the above we may conclude that

$$|\varepsilon_3|\le C\tau\to 0. \tag{7.48}$$

Recall the definition of $T_{\ell\tau}$ in (7.33). Denote the analogues of $T_{\ell\tau}$ by T_ℓ with the truncated matrix $\mathbf{X}(s)$ replaced by the initial matrix $\mathbf{X}(s)$. Then write

$$T_{\ell\tau}=T_\ell+r_\ell,\quad \ell=0,1,2,3,$$

where

$$\begin{aligned}|r_\ell| \;\le\; & \frac{s^{(\ell-1)/2}}{\ell! n^{(\ell+1)/2}}\sum_{j,t=1}^{p,n}|\kappa_{(\ell+1),\tau}-\kappa_{\ell+1}|\Big|\sum_{k=0}^{\infty}|b_k| \\ & \sum_{k^{(\ell)},k^{(\ell-1)},\ldots,k^{(1)}}^{p}|b_{k^{(\ell)}-j+k}b_{k^{(\ell-1)}-j+k}\cdots b_{k^{(1)}-j+k}|E\Big(D_{k^{(\ell)}t}D_{k^{(\ell-1)}t}\cdots D_{k^{(1)}t}\Psi_{jt}(s)\Big)\Big| \\ \le\; & \frac{C}{(\tau)^{3-\ell}}E\Big(|\xi_{11}|^4\cdot I(|\xi_{11}|>\tau\sqrt{n})\Big)\to 0,\end{aligned}$$

where the last step uses (7.29), (7.41) and an estimate similar to (7.21).

By Lemma 6 below, (7.33), (7.48) and the facts that $T_0=T_3=0$ (because $\kappa_1=\kappa_4=0$) and that $T_1=B_n$ (see (7.32)) we see

$$D_n=B_n+o(1).$$

This, together with (7.31), ensures (7.30) by the facts that $|a_n|=1$ and that the function f is an analytic function. The proof of theorem is complete. □

Lemma 6.

$$T_2=\frac{s^{1/2}\kappa_3}{2n^{3/2}}\sum_{j,t=1}^{p,n}\sum_{k=0}^{\infty}b_k\sum_{k^{(2)},k^{(1)}=j-k}^{p}b_{k^{(2)}-j+k}b_{k^{(1)}-j+k}E\Big(D_{k^{(2)}t}(s)D_{k^{(1)}t}(s)\Psi_{jt}(s)\Big)=o(1),$$

as $n\to\infty$.

Proof. It follows from (7.35)-(7.38) that the expansion of $D_{k^{(1)}t}\Psi_{jt}(s)$ is

$$\begin{aligned}D_{k^{(1)}t}\Psi_{jt}(s) \;=\; & e_n(s,x)\Big[-2x\int\theta\hat{f}(\theta)\big(\mathbf{UY}(s)\big)_{k^{(1)}t}d\theta\big(\mathbf{UY}(s)\big)_{jt}+U_{k^{(1)}j} \\ & +i\big(\mathbf{Y}'(s)\mathbf{UY}(s)\big)_{tt}*U_{k^{(1)}j}+i\big(\mathbf{UY}(s)\big)_{jt}*\big(\mathbf{UY}(s)\big)_{k^{(1)}t}\Big].\end{aligned} \tag{7.49}$$

By (7.35)-(7.38) we can further obtain the expansion of $D_{k^{(2)}t}D_{k^{(1)}t}\Psi_{jt}(s)$. Since such an expansion is complicated we do not list it here. However each term of the expansion is a constant multiple of one of the following forms

$$A_1 = \big(\mathbf{UY}(s)\big)_{h_1t} \circ U_{h_2h_3}e_n(s,x),$$
$$A_2 = \big(\mathbf{UY}(s)\big)_{h_1t} \circ \big(\mathbf{Y}^T(s)\mathbf{UY}(s)\big)_{tt} \circ U_{h_2h_3}e_n(s,x),$$
$$A_3 = \big(\mathbf{UY}(s)\big)_{k^{(2)}t} \circ \big(\mathbf{UY}(s)\big)_{k^{(1)}t} \circ \big(\mathbf{UY}(s)\big)_{jt}e_n(s,x),$$

where "$\circ$" denotes a product or a convolution; $h_i = k^{(2)}, k^{(1)}$ or j with $i = 1,2,3$ and $h_1 \neq h_2 \neq h_3$. In view of this it then suffices to prove that

$$T_{2i} = \frac{1}{n^{3/2}}\sum_{j,t=1}^{p,n}\sum_{k=0}^{\infty} b_k \sum_{k^{(1)},k^{(2)}=j-k}^{p} b_{k^{(1)}-j+k}b_{k^{(2)}-j+k}EA_i = o(1), \quad i = 1,2,3.$$

Without loss of generality, we below consider $h_1 = j$, $h_2 = k^{(1)}$ and $h_3 = k^{(2)}$ only, otherwise one may first exchange the order of the summation as in (7.47) when necessary and then proceed as follows. Consider T_{22}. Note that the fact that $\mathbf{UY}(s)\mathbf{Y}^T(s) = \mathbf{Y}(s)\mathbf{Y}^T(s)\mathbf{U}$. A simple calculation then yields

$$E\Big[\sum_{t=1}^{n}|\sum_{j=1}^{p}\big(\mathbf{UY}(s)\big)_{jt}|^2\Big] = E\Big[\sum_{j_1,j_2=1}^{p}\big(\mathbf{UY}(s)\mathbf{Y}^T(s)\bar{\mathbf{U}}^T\big)_{j_1j_2}\Big]$$
$$= E\Big[\sum_{j_1,j_2=1}^{p}\big(\mathbf{Y}(s)\mathbf{Y}^T(s)\big)_{j_1j_2}\Big] = O(n). \tag{7.50}$$

By the Schwartz inequality, (7.39) and (7.50), we have

$$|T_{22}|^2 \leq \frac{C}{n^3}E\Big[\sum_{t=1}^{n}|\big(\mathbf{Y}^T(s)\mathbf{UY}(s)\big)_{tt}|^2\Big]E\Big[\sum_{t=1}^{n}|\sum_{j=1}^{p}\big(\mathbf{UY}(s)\big)_{jt}|^2\Big]$$
$$\leq \frac{C}{n^2}E\Big[\sum_{t=1}^{n}\big(\sum_{r=1}^{p}Y_{rt}^2(s)\big)^2\Big] = O(\frac{1}{n}). \tag{7.51}$$

This argument also works for T_{21} and T_{23} and we ignore the details here. Therefore

$$T_2 = O(\frac{1}{\sqrt{n}}).$$

□

Proof of Theorem 4. Set

$$X_n^{(i)} = \int x^i d\widetilde{G}_n(x), \quad i = 1,2; \qquad \widetilde{\mathbf{\Omega}}^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}.$$

Furthermore, under $\mathbb{H}_1$, $X_n^{(i)}$, $i = 1, 2$, can be written as

$$X_n^{(i)} \quad = \quad Y_1^{(i)} + Y_2^{(i)}, \tag{7.52}$$

where

$$Y_1^{(i)} = p\int x^i d\Big(F_{\mathbb{H}_1}^{\mathbf{S}^{(1)}}(x) - F_{\mathbb{H}_0}^{\mathbf{S}^{(1)}}(x)\Big) + p\int x^i d\Big(F_{\mathbb{H}_0}^{\mathbf{S}^{(2)}}(x) - F_{\mathbb{H}_1}^{\mathbf{S}^{(2)}}(x)\Big)$$

and

$$Y_2^{(i)} = p\int x^i d\Big(F_{\mathbb{H}_0}^{\mathbf{S}^{(1)}}(x) - F_{\mathbb{H}_0}^{\mathbf{S}^{(2)}}(x)\Big).$$

From (7.52) we have

$$a_{11}(X_n^{(1)})^2 + a_{22}(X_n^{(2)})^2 + 2a_{12}X_n^{(1)}X_n^{(2)} = W_1 + W_2 + W_3,$$

where

$$W_1 = a_{11}(Y_2^{(1)})^2 + a_{22}(Y_2^{(2)})^2 + 2a_{12}Y_2^{(1)}Y_2^{(2)},\ W_2 = a_{11}(Y_1^{(1)})^2 + a_{22}(Y_1^{(2)})^2 + 2a_{12}Y_1^{(1)}Y_1^{(2)}$$

and

$$W_3 = 2a_{11}Y_1^{(1)}Y_2^{(1)} + 2a_{22}Y_1^{(2)}Y_2^{(2)} + 2a_{12}[Y_2^{(1)}Y_1^{(2)} + Y_1^{(1)}Y_2^{(2)}].$$

Note that W_1 converges in distribution to $\chi^2(2)$ by Theorem 2 or Proposition 1. Also $Y_2^{(i)}, i = 1, 2$ converge in distribution to Gaussian distribution by Theorem 2 or Proposition 1. We next prove that $W_2 \to \infty$ in probability while $W_3 = o_p(W_2)$. By Assumption (3.15) $Y_1^{(1)} \to \infty$ or $Y_1^{(2)} \to \infty$ in probability (we would point out that $Y_1^{(i)} \geq 0$). If $Y_1^{(1)} \to \infty$ and $\limsup Y_1^{(2)} < \infty$ in probability, then $W_2 \to +\infty$ in probability. It is then easy to verify that $W_3 = o_p(W_2)$. This argument also applies to the case when $Y_1^{(2)} \to \infty$ and $\limsup Y_1^{(1)} < \infty$ in probability. If $Y_1^{(1)} \to \infty$ and $Y_1^{(2)} \to \infty$ in probability then by Holder's inequality

$$W_2 \geq 2(\sqrt{a_{11}}\sqrt{a_{22}} + a_{12})Y_1^{(1)}Y_1^{(2)} \to +\infty$$

in probability, because

$$det(\widetilde{\mathbf{\Omega}}^{-1}) = a_{11}a_{22} - a_{12}^2 > 0.$$

It is then easy to verify that $W_3 = o_p(W_2)$ in this case.

In view of the above we conclude from the definition of L_n that

$$\begin{aligned} P(L_n > \gamma_{1-\alpha}|\mathbb{H}_1) &= P\Big((X_n^{(1)}, X_n^{(2)})\widetilde{\mathbf{\Omega}}^{-1}\begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \end{pmatrix} > \gamma_{1-\alpha}\Big|\mathbb{H}_1\Big) \\ &= P\Big((a_{11}(X_n^{(1)})^2 + a_{22}(X_n^{(2)})^2 + 2a_{12}X_n^{(1)}X_n^{(2)} > \gamma_{1-\alpha}\Big|\mathbb{H}_1\Big) \\ &= P\Big(W_1 + W_2 + W_3 > \sqrt{\gamma_{1-\alpha}}\Big|\mathbb{H}_1\Big) \to 1, \quad as \quad n \to \infty. \end{aligned} \tag{7.53}$$

□

References

Anderson, T. W. (1984). An introduction to multivariate statistical analysis. *New York: Wiley* **2th edtion**.

Bai, Z. D. and Silverstein, J. (2009). Spectral analysis of large dimensional random matrices. *Science Press* **2nd edition**.

Birke, M. and Dette, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. and Probab. Letters.* **74**, 281-289.

Chen, J., Gao, J. and Li, D. (2012). A new diagnostic test for cross–sectional independence in nonparametric panel data models. *Econometric Theory* **28**, 1144–1163.

Frees, E. W. (1995). Assessing cross-sectional correlation in panel data. *Journal of Econometrics* **69**, 393-414.

Geweke, J. (1981). A comparison of tests of independence of two covariance stationary time series. *J. Am. Statist. Assoc.* **76**, 363-373.

Haugh, L. D. (1976). Checking the independence of two covariance stationary time series: a univariate residual cross correlation approach. *J. Am. Statist. Assoc.* **71**, 378-385.

Hong,Y.M. (1996). Testing for independence between two covariance stationary time series. *Biometrika* **83(3)**, 615-625.

Hong,Y.M. (1999). Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *J. Am. Statist. Assoc.* **94(448)**, 1201-1220.

Hong,Y. M. (2000). Generalized spectral tests for serial dependence. *J. R. Statist. Soc. B* **62(3)**, 557-574.

Hong,Y. M. (2005). Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* **73(3)**, 837-901.

Hsiao, C., Pesaran, M. H. and Pick, A. (2009). Diagnostic tests of cross–sectional independence for limited dependent variable panel data models. *Oxford Bulletin of Economics and Statistics.*

Johnstone, I. (2001). on the distribution of the largest principal component. *Ann. Statist.* **29**, 295-327.

Ledoit, O and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compare to the sample size. *Ann. Statist.* **30**, 1081-1102.

Lytova, A. and Pastur, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37(5)**, 1778-1840.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica* **46**, 69-85.

Pan, G. M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18(3)**, 1232-1270.

Rama, C. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quntitative Finance* **1**, 223-236.

Gray, R. M. (2009). Toeplitz and Circulant Matrices: A review. http://ee.stanford.edu/ gray/toeplitz.pdf.

Sarafidis,V., Yamagata, T. and Robertson, D. (2009). A test of cross section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics* **148(2)**, 149-161.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices, *J. Multivariate Anal.* **55**(1995), 331–339.

Yao, J. F. (2012). A note on a MarcenkoPastur type theorem for time series. *Statist. and Probab. Letters.* **82(1)**, 22-28.

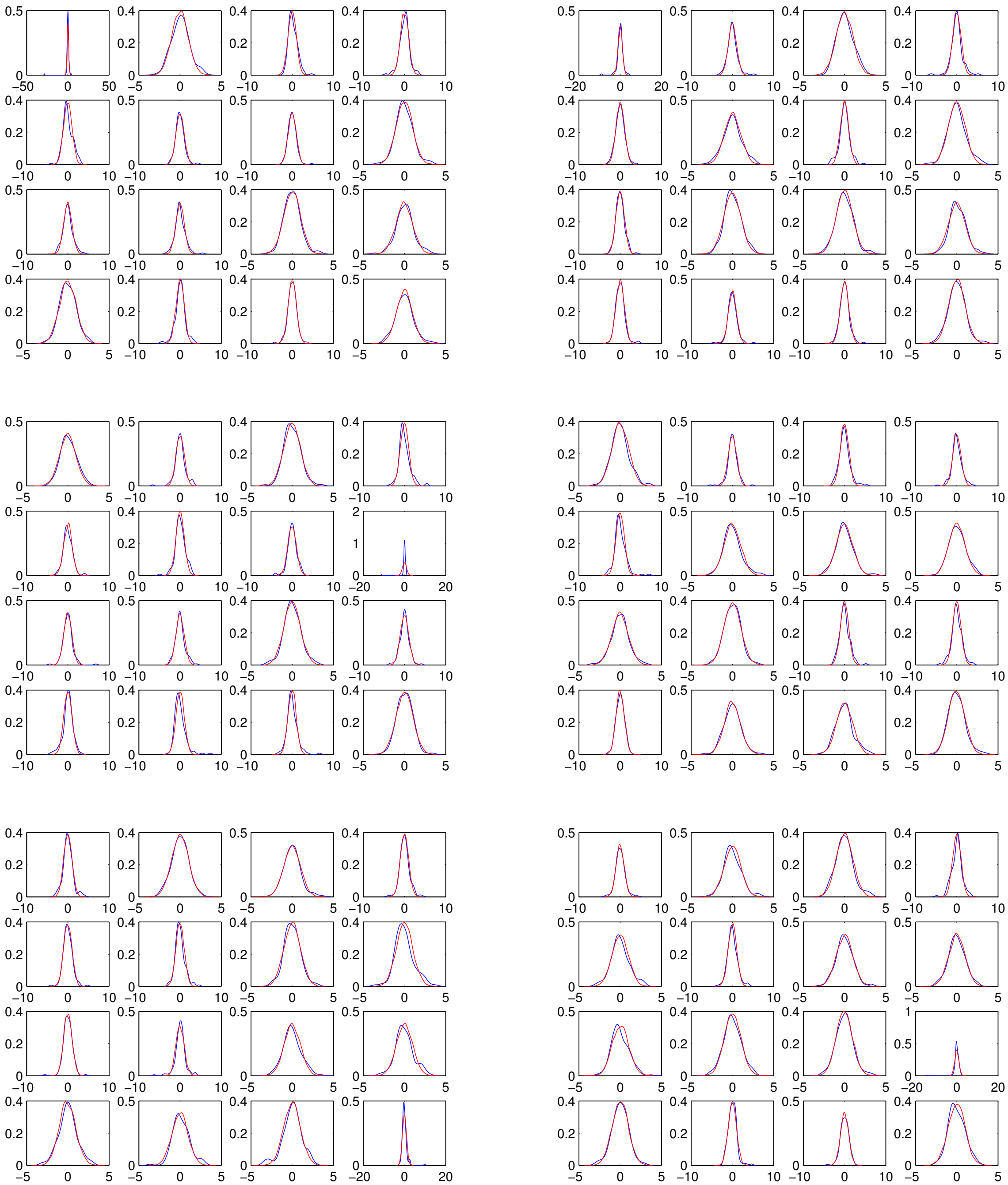
Zhang, L. X. (2006). Spectral analysis of large dimensional random matrices. *National University of Singapore PHD Thesis.*

Zheng, S. R. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. H. Poincare Probab. Statist.* **48(2)**, 444-476.

Table 1: Empirical sizes of the proposed test L_n at significant level 0.05 for n time series generated from DGP 1 with $\theta_1 = 0.8$ in model (4.3).

	p									
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.023	0.022	0.027	0.021	0.019	0.021	0.023	0.018	0.022	0.020
100	0.037	0.040	0.035	0.039	0.036	0.031	0.034	0.030	0.029	0.025
150	0.042	0.040	0.039	0.043	0.041	0.038	0.039	0.034	0.034	0.031
200	0.039	0.043	0.046	0.045	0.043	0.045	0.042	0.041	0.038	0.040
250	0.040	0.045	0.048	0.044	0.045	0.044	0.042	0.040	0.041	0.041
300	0.037	0.041	0.049	0.054	0.052	0.048	0.043	0.041	0.039	0.045
350	0.041	0.048	0.053	0.049	0.055	0.052	0.049	0.047	0.045	0.047
400	0.038	0.041	0.047	0.052	0.052	0.048	0.053	0.051	0.046	0.046
450	0.035	0.037	0.040	0.046	0.050	0.055	0.059	0.060	0.058	0.055
500	0.032	0.035	0.035	0.040	0.047	0.052	0.054	0.057	0.054	0.058

Figure 1: Graphs of smoothed density function of the transformed data vs standard normal distribution



*These graphs contain the empirical density functions of the transformed data for all 96 stocks used in our empirical application. The blue line is the smoothed density function of the transformed data for one stock and the red graph is standard normal density function.

Table 2: Bootstrap sizes of the proposed test L_n at significant level 0.05 for n time series generated from DGP 1 with $\theta_1 = 0.8$ in model (4.3).

	p									
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.039	0.041	0.038	0.037	0.040	0.037	0.039	0.035	0.037	0.037
100	0.043	0.042	0.040	0.039	0.039	0.040	0.042	0.040	0.046	0.041
150	0.046	0.042	0.048	0.043	0.045	0.047	0.040	0.049	0.045	0.041
200	0.042	0.047	0.043	0.046	0.049	0.051	0.048	0.048	0.046	0.050
250	0.052	0.050	0.046	0.054	0.048	0.051	0.050	0.049	0.052	0.055
300	0.048	0.053	0.055	0.052	0.056	0.057	0.053	0.051	0.049	0.054
350	0.046	0.054	0.052	0.054	0.050	0.051	0.050	0.048	0.053	0.054
400	0.043	0.048	0.046	0.051	0.054	0.051	0.052	0.054	0.049	0.052
450	0.046	0.052	0.048	0.049	0.053	0.050	0.054	0.055	0.053	0.052
500	0.042	0.046	0.045	0.047	0.050	0.053	0.055	0.052	0.055	0.054

Table 3: Empirical sizes of the proposed test L_n at significant level 0.05 for n time series generated from DGP 2.

					p					
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.037	0.035	0.030	0.031	0.034	0.029	0.032	0.030	0.030	0.028
100	0.040	0.043	0.045	0.042	0.040	0.037	0.040	0.035	0.035	0.032
150	0.042	0.043	0.048	0.043	0.045	0.040	0.040	0.039	0.037	0.037
200	0.041	0.046	0.051	0.045	0.049	0.053	0.047	0.043	0.040	0.041
250	0.044	0.049	0.053	0.051	0.047	0.052	0.044	0.047	0.045	0.045
300	0.040	0.042	0.046	0.046	0.050	0.055	0.047	0.044	0.046	0.048
350	0.038	0.046	0.049	0.053	0.054	0.051	0.053	0.046	0.045	0.046
400	0.039	0.041	0.043	0.047	0.055	0.051	0.058	0.056	0.051	0.053
450	0.037	0.039	0.039	0.043	0.049	0.053	0.055	0.048	0.050	0.048
500	0.037	0.035	0.042	0.047	0.045	0.055	0.047	0.054	0.052	0.055

*The data are simulated from model (4.4). $\phi_1 = 0.2$.

Table 4: Bootstrap sizes of the proposed test L_n at significant level 0.05 for n time series generated from DGP 2.

	p									
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.041	0.039	0.040	0.042	0.038	0.037	0.035	0.038	0.036	0.038
100	0.044	0.047	0.043	0.040	0.042	0.041	0.040	0.045	0.039	0.040
150	0.046	0.048	0.049	0.051	0.048	0.047	0.053	0.055	0.052	0.047
200	0.043	0.050	0.053	0.049	0.052	0.048	0.049	0.054	0.051	0.047
250	0.045	0.052	0.054	0.050	0.050	0.054	0.051	0.048	0.047	0.052
300	0.044	0.048	0.051	0.047	0.048	0.050	0.046	0.053	0.052	0.047
350	0.046	0.051	0.047	0.054	0.052	0.050	0.051	0.051	0.052	0.049
400	0.042	0.047	0.052	0.049	0.051	0.050	0.055	0.050	0.054	0.052
450	0.045	0.049	0.053	0.053	0.050	0.051	0.054	0.049	0.049	0.051
500	0.042	0.035	0.045	0.049	0.047	0.050	0.051	0.047	0.050	0.053

*The data are simulated from model (4.4). $\phi_1 = 0.2$.

Table 5: Empirical sizes of the proposed test L_n at significant level 0.05 for n time series generated from DGP 3.

	p									
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.036	0.039	0.035	0.030	0.035	0.031	0.036	0.035	0.030	0.029
100	0.040	0.042	0.040	0.039	0.042	0.045	0.040	0.037	0.039	0.034
150	0.043	0.048	0.051	0.045	0.047	0.051	0.043	0.042	0.039	0.040
200	0.040	0.046	0.055	0.052	0.047	0.055	0.049	0.045	0.044	0.044
250	0.039	0.041	0.048	0.053	0.051	0.057	0.053	0.055	0.058	0.059
300	0.042	0.045	0.045	0.050	0.048	0.055	0.054	0.048	0.052	0.056
350	0.037	0.042	0.047	0.052	0.050	0.051	0.047	0.045	0.048	0.052
400	0.035	0.045	0.052	0.047	0.051	0.048	0.053	0.054	0.052	0.050
450	0.038	0.041	0.045	0.046	0.045	0.047	0.049	0.052	0.050	0.048
500	0.039	0.043	0.047	0.052	0.048	0.048	0.051	0.055	0.046	0.051

*The data are simulated from model (4.5). $\theta_1 = 0.8$ and $\phi_1 = 0.2$.

Table 6: Bootstrap sizes of the proposed test L_n at significant level 0.05 for n time series generated from DGP 3.

					p					
n	50	100	150	200	250	300	350	400	450	500
					Empirical sizes					
50	0.040	0.042	0.045	0.037	0.035	0.034	0.032	0.036	0.034	0.038
100	0.043	0.045	0.048	0.046	0.039	0.042	0.046	0.047	0.044	0.044
150	0.047	0.052	0.049	0.055	0.057	0.050	0.046	0.045	0.040	0.043
200	0.045	0.048	0.053	0.050	0.051	0.052	0.055	0.055	0.056	0.052
250	0.047	0.048	0.051	0.055	0.054	0.056	0.056	0.052	0.054	0.053
300	0.045	0.049	0.050	0.053	0.050	0.053	0.058	0.054	0.053	0.055
350	0.048	0.053	0.057	0.053	0.052	0.050	0.049	0.047	0.053	0.055
400	0.045	0.048	0.050	0.055	0.051	0.054	0.053	0.056	0.055	0.056
450	0.042	0.044	0.051	0.048	0.053	0.054	0.050	0.051	0.053	0.053
500	0.045	0.048	0.050	0.054	0.050	0.049	0.050	0.053	0.053	0.054

*The data are simulated from model (4.5). $\theta_1 = 0.8$ and $\phi_1 = 0.2$.

Table 7: Empirical powers of the proposed test L_n at significant level 0.05 for n time series with MA(1) type dependent structure.

			p			
n	50	100	200	300	350	400
			Empirical sizes			
50	0.210	0.279	0.429	0.445	0.505	0.614
100	0.469	0.513	0.725	0.779	0.794	0.805
200	0.712	0.793	0.814	0.889	0.903	0.921
300	0.787	0.899	0.932	0.921	0.945	0.962
350	0.823	0.956	0.983	0.972	0.989	0.994
400	0.921	0.993	0.994	0.999	1.000.	0.999

*The data are simulated from model (4.6). Each time series $\mathbf{x}_i$ is generated from DGP 3 with $\theta_1 = 0.8$ and $\phi_1 = 0.2$. In (4.6), we take $\theta = 0.8$.

Table 8: Empirical powers of the proposed test L_n at significant level 0.05 for n time series with AR(1) type dependent structure.

			p			
n	50	100	200	300	350	400
			Empirical sizes			
50	0.656	0.720	0.714	0.801	0.823	0.842
100	0.792	0.824	0.846	0.891	0.907	0.917
200	0.858	0.889	0.922	0.926	0.954	0.985
300	0.901	0.935	0.958	0.982	0.992	0.0.993
350	0.892.	0.970	0.992	0.995	0.999	0.999
400	0.941	0.989	0.999	1.000	1.000	1.000

*The data are simulated from model (4.7). Each time series $\mathbf{x}_i$ is generated from DGP 3 with $\theta_1 = 0.8$ and $\phi_1 = 0.2$. In (4.7), we take $\phi = 0.2$.

Table 9: Empirical powers of the proposed test L_n at significant level 0.05 for n time series with ARMA(1,1) type dependent structure.

			p			
n	50	100	200	300	350	400
			Empirical sizes			
50	0.592	0.613	0.654	0.719	0.746	0.758
100	0.713	0.748	0.855	0.891	0.904	0.909
200	0.776	0.833	0.892	0.903	0.955	0.968
300	0.856	0.901	0.963	0.981	0.982	0.993
350	0.902	0.946	0.980	0.999	0.998	1.000
400	0.933	0.951	0.991	1.000	1.000	1.000

*The data are simulated from model (4.8). Each time series $\mathbf{x}_i$ is generated from DGP 3 with $\theta_1 = 0.8$ and $\phi_1 = 0.2$. In (4.8), we take $\theta = 0.8$ and $\phi = 0.2$.

Table 10: Empirical powers of the proposed test L_n at 0.05 significance level for the dynamic factor model.

(p, n)	r=1	r=2	r=3	r=4
(50,50)	0.342	0.553	0.889	0.950
(50,100)	0.358	0.622	0.949	0.968
(100,100)	0.403	0.685	0.972	0.984
(200,100)	0.526	0.741	0.983	0.998
(300,200)	0.557	0.763	0.987	1.000
(200,300)	0.637	0.785	0.983	0.999
(100,200)	0.656	0.791	0.988	0.999
(200,400)	0.671	0.785	0.990	0.999
(400,200)	0.685	0.768	0.991	1.000
(100,300)	0.682	0.784	0.980	1.000
(300,100)	0.701	0.782	0.989	1.000

*The data are simulated from model (4.9) and (4.10).

Table 11: Empirical powers of the proposed test L_n at significant level 0.05 for n random vectors with common random dependence.

	p					
n	50	70	90	110	130	150
			Empirical sizes			
50	0.894	0.920	0.923	0.942	0.966	0.959
70	0.910	0.948	0.955	0.975	0.980	0.995
90	0.960	0.958	0.969	0.984	0.989	0.999
110	0.941	0.956	0.984	0.992	0.994	1.000
130	0.930	0.972	0.990	0.995	0.999	1.000
150	0.952	0.980	0.989	1.000	1.000	1.000

*The data are simulated from model (4.11).

Table 12: Empirical powers of the proposed test L_n at 0.05 significance level for ARCH(1) dependent type.

(p, n)	$(0.9, 0.1)$	$(0.8, 0.2)$	$(0.7, 0.3)$	$(0.6, 0.4)$	$(0.5, 0.5)$
(50,50)	0.257	0.396	0.425	0.605	0.732
(50,100)	0.597	0.879	0.890	0.899	0.998
(100,200)	0.727	0.978	0.997	0.998	0.999
(200,200)	0.738	0.990	0.999	1.000	1.000
(200,300)	0.828	0.992	0.998	1.000	1.000
(200,400)	0.887	0.997	1.000	1.000	1.000
(300,400)	0.906	1.000	1.000	1.000	1.000
(400,400)	0.922	1.000	1.000	1.000	1.000

*The data are simulated from model (4.12).

Table 13: L_n under various scenarios for 5 randomly selected samples

(n,p)	5%critical values	1	2	3	4	5
(60,30)	[0, 5.99]	395.44	462.76	481.85	443.79	481.46
(70,35)	[0, 5.99]	595.84	642.31	620.96	592.63	632.87
(90,40)	[0, 5.99]	902.55	928.89	1318.6	1173.9	914.25

*The critical values are the corresponding quantiles of the limiting distribution $\chi^2(2)$ of the statistic L_n for $(n, p) = (60, 30), (70, 35), (90, 40)$ respectively.