Detecting linear and nonlinear dependence for a large number of random variables

Zhigang Bao
Department of Mathematics, Zhejiang University
Guangming Pan, Yanrong Yang
Division of Mathematical Sciences, Nanyang Technological University
Wang Zhou
Department of Statistics and Applied Probability, National University of Singapore

Abstract

This paper considers measuring dependence between the components of a random vector where the dimensionality of the random vector is comparable to the number of observations. It is well known that the standardized population covariance matrix can describe pairwise dependence of a Gaussian random vector. However, for non-Gaussian data, this criteria may be failed with missing nonlinear dependent relationships with zero correlations. We propose a new way of detecting dependence of the investigated data by considering the population covariance matrix of a transformed data which takes into account high order correlations between the variables. Another difficulty in tackling this problem is that high dimensionality of the data brings in many complications, such as the role of a consistent estimator played by the corresponding sample covariance matrix disappears. This point hinders utilizing the intuition that the integral properties of the sample covariance matrix tend to be similar with those of its corresponding population covariance matrix. We study the asymptotic trend of linear spectral statistics of the sample covariance matrix for the transformed data under independent case. Relying on this characterization, it can be used as a tool of discriminating between dependence and independence. Except of the capability of catching nonlinear dependence, another key advantage lies on the fact that this method does not require any moment condition on the data and thus it can be applied to heavy tailed data. Its good practical performance is demonstrated in an extensive simulation study. Finally, an empirical application to stock returns of S&P500 also illustrates its effectiveness.

Keywords: Central limit theorem; Generalized sample covariance matrices; High dimensional data; Hypothesis test; Linear spectral statistics; Nonlinear dependence; Random matrices; Stieltjes transform.

JEL Classifications: C12, C15, C18.

1 Introduction

Inferring the relationship among random variables, for example, the relationship between the response and the explanatory variables in linear models, is an extremely important and widely studied statistical problem from the practical and theoretical points of view. This paper considers detecting dependence among a large number of random variables.

Recent technological innovations have brought explosion of data into many scientific disciplines, including genomics, image processing, microarray, proteomics and finance, to name but a few. The dimensionality of the data p can be much larger than or comparable to the sample size n. We focus on the scenario of p/n tending to a constant.

Consider a p-dimensional random vector(or a time series with length p) $\mathbf{x} = (X_1, X_2, \ldots, X_p) \in \mathbb{R}^p$ and let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be a random sample of size n from the population. Dependence of the p components in $\mathbf{x}$ is often characterized by the population covariance matrix $\mathbf{\Sigma}$ of $\mathbf{x}$. Any deviation of the matrix $\mathbf{\Sigma}$ from a diagonal matrix provides evidence of dependence. This idea can be adopted to capture all types of dependencies for a set of Gaussian variables. It is natural to focus on the sample covariance matrix $\mathbf{S}$ to develop independence test statistics for normal distributed data. However, in high dimensional data analysis, the dimension p and the sample size n are both large, and p may be comparable to n or even much larger than n. The phenomenon called 'curse of dimensionality' appears, i.e. the sample covariance matrix $\mathbf{S}$ is not a consistent estimator of the population covariance matrix $\mathbf{\Sigma}$. However it is still possible to derive accurate information on $\mathbf{\Sigma}$ from $\mathbf{S}$. Recently, there is a long list of literature devoted to investigating properties of the large dimensional sample covariance matrix $\mathbf{S}$ under the null hypothesis that $\mathbf{\Sigma} = \mathbf{I}_p$. For example, when $p/n \to c \in (0, \infty)$, Johnstone (2001) and Bao, Pan and Zhou (2012) established the Tracy-Widom law of the largest eigenvalue of the sample covariance matrix and correlation matrix respectively; Ledoit and Wolf (2002) studied the quadratic form of the eigenvalues of $\mathbf{S}$; Schott (2005) discussed the sums of squares of the sample correlation coefficients; Cai and Jiang (2011) considered the largest entry of the sample correlation matrix and Liu, Lin and Shao (2008) introduced a modified test statistic based on the largest entry of the sample correlation matrix; Bai et al. (2009) proposed a correction to the likelihood ratio test statistic which is a linear spectral statistic of the sample covariance matrix.

However, for non-Gaussian case, the methodology of comparing the population covariance matrix $\mathbf{\Sigma}$ with an identity matrix is invalid in dealing with a random vector whose elements possess zero correlations but display nonlinear dependence[1]. Such examples appearing in nonlinear time series include autoregressive conditional heteroscedastic (ARCH), bilinear and nonlinear moving average processes, etc.; See Tong (1990) for a detailed review. In the recent works Bao (2018), Bao (2019) and Bao, Lin, Pan and Zhou (2015), the spectral statistics of the nonparametric matrices such as the Spearman correlation matrix and Kendall correlation matrices are studied for the detection of nonlinear independence. Besides, Hong (1999) proposed a generalized spectral density approach to capturing the nonlinear dependence, which is suitable for both linear and nonlinear strictly stationary processes. Surprisingly there is no other research on this paper.

Considering that high-order moments of $X_1, X_2, \ldots, X_p$ can characterize their nonlinear dependence, we suggest to study the covariance matrix $\mathbf{\Sigma}_e$ of the modified random vector $\mathbf{e}^{it\mathbf{x}} := (e^{itX_1}, e^{itX_2}, \ldots, e^{itX_p})$ with any real constant t since $Cov(e^{itX_j}, e^{itX_k})$ is able to capture all kinds of dependence between X_j and X_k, including those with zero correlations [2]. Therefore a new statistic based on the empirical spectral distribution of the sample covariance matrix $\mathbf{S}_e$ of the vector $\mathbf{e}^{it\mathbf{x}}$ is proposed in this paper. In other words, the dependent relationship among the initial variables in $\mathbf{x}$ is reflected by the correlations among the components of $\mathbf{e}^{it\mathbf{x}}$.

One key ingredient of our methodology, which sets it apart from other approaches listed above, is that there is no moment condition on the original data. This advantage engages

[1]Though some existing statistics are proposed under independent assumption, simulation results show that these ones fail for uncorrelated but dependent cases.

[2]$Cov(e^{iuX_j}, e^{ivX_k}) = 0$ for all $(u, v) \in \mathbb{R}^2$ if and only if X_j and X_k are independent(see Lukacs (1970)). X_j and X_k are called subindependent if $Cov(e^{itX_j}, e^{itX_k}) = 0$ for any $t \in \mathbb{R}$, which is equivalent to $\sum_{s+r=m} Cov(X_j^s, X_k^r) = 0$ for any $m \in \mathbb{N}^+$.

that our method can be applied to the heavy-tailed distribution which provides a good fit to the financial data. In addition, most of the data in economics, geology, climatology, signal processing, insurance, environmetrics do have heavy-tailed distributions. See Adler et al. (1998).

Moreover, as an contribution to the large dimensional random matrix theory, the asymptotic theorem provided in this paper generalizes the central limit theorem for large dimensional sample covariance matrices in the complex case(see Bai and Silverstein (2004)) by eliminating the condition that $EX_{ij}^2 = 0$.

The remainder of this article is organized as follows. Section 2 introduces some basic concepts and main results for large dimensional sample covariance matrices. In Section 3, we propose a new dependence test statistic for a large dimensional random vector. Moreover, the asymptotic theory for the proposed statistic is presented. Simulation results are provided in Section 4. Some classical dependent structures appearing in time series are illustrated there. An empirical application to the stock returns from S&P 500 is given in Section 5. Section 6 concludes the paper and the Appendix contains the outline of proofs. Detailed proofs are included in a separate supplementary material.

2 Preliminary

Let $X_1, X_2, \ldots, X_p$ be p random variables with identical distribution $\mathbf{F}$. Group these p random variables into a random vector $\mathbf{x} = (X_1, X_2, \ldots, X_p)$ and suppose a random sample from $\mathbf{x}$ of size n, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, with $\mathbf{x}_k = (X_{1k}, X_{2k}, \ldots, X_{pk})'$, $k = 1, 2, \ldots, n$. The goal is to construct a statistic to test the hypothesis that

$$H_0 : X_1, X_2, \ldots, X_p \text{ are independent}; \text{ against } H_1 : X_1, X_2, \ldots, X_p \text{ are dependent.} \quad (2.1)$$

Consider an alternative sample covariance matrix $\mathbf{S}_e = \frac{1}{n}\mathbf{Y}\mathbf{Y}^*$, where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$ and $\mathbf{y}_k = (e^{itX_{1k}}, e^{itX_{2k}}, \ldots, e^{itX_{pk}})'$, $k = 1, 2, \ldots, n$, with $\mu = Ee^{itX_{11}}$ and $\sigma = Var(e^{itX_{11}})$. Relying on the idea of converting dependence among the p random variables in $\mathbf{x}$ into correlations among the components of $\mathbf{y}$, a new statistic based on the empirical spectral distribution (ESD) of the matrix $\mathbf{S}_e$ is proposed. The ESD of a $p \times p$ matrix $\mathbf{A}$ is defined as

$$F^{\mathbf{A}}(x) = \frac{1}{p}\#\{j : \lambda_j \leq x, \ 1 \leq j \leq p\},$$

where $\#\{\cdots\}$ denotes the cardinality of the set $\{\cdots\}$ and $\lambda_j, \ j = 1, 2, \ldots, p$ are eigenvalues of the matrix $\mathbf{A}$. Under the null hypothesis H_0, it is well known that the limiting spectral distribution (LSD) of the sample covariance matrix $\mathbf{S}_e$ is the M-P (Marcĕnko and Pastur) law whose density function is

$$p_c(x) = \begin{cases} \frac{1}{2\pi xc}\sqrt{(b-x)(x-a)}, & a \leq x \leq b, \\ 0, & otherwise, \end{cases}$$

and has a point mass $1 - 1/c$ at the origin if $c > 1$, where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$. Here $c = \lim_{n\to\infty} p/n$. Another important concept in random matrix theory is the Stieltjes transform. The Stieltjes transform of any probability distribution function $G(x)$ is defined as

$$s_G(z) = \int \frac{1}{x-z} dG(x), \ z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C}, v = \Im z > 0\}.$$

When $X_1, X_2, \ldots, X_p$ are not independent, $e^{itX_1}, e^{itX_2}, \ldots, e^{itX_p}$ are correlated for some $t > 0$. Bai and Zhou (2008) provided the LSD of sample covariance matrices with correlated structures. For easy reference, we state it in the following proposition:

Proposition 1. *As $n \to \infty$, assume the following.*
1. For all k, $E\bar{Y}_{\ell k}Y_{jk} = t_{\ell j}$, and for any non-random $p \times p$ matrix $\mathbf{B} = (b_{j\ell})$ with bounded norm, $E|\mathbf{y}_k^*\mathbf{B}\mathbf{y}_k - tr\mathbf{B}\mathbf{T}| = o(n^2)$, where $\mathbf{T} = (t_{j\ell})_{p\times p}$.
2. $c_n = p/n \to c \in (0, +\infty)$.
3. The norm of the matrix $\mathbf{T}$ is uniformly bounded and its $F^{\mathbf{T}}$ tends to a non-random probability distribution H.
Then with probability one, $F^{\mathbf{S}_e}$ tends to a probability distribution, whose Stieltjes transform $s(z), z \in \mathbb{C}$ satisfies

$$s(z) = \int \frac{1}{t(1 - c - czs(z)) - z} dH(t). \tag{2.2}$$

From (2.2), we can see that the LSD $F^{\mathbf{S}_e}$ of sample covariance matrices depends on the correlation structure, i.e. $H(t)$, of the transformed data $\mathbf{y}$. When $\mathbf{T} = \mathbf{I}_p$, the identity matrix, $F^{\mathbf{S}_e}$ tends to the M-P law, whose Stieltjes transform is

$$s(z) = \frac{1 - c - z + \sqrt{(1 + c - z)^2 - 4c}}{2cz}.$$

So the difference of the LSD of the sample covariance matrix $\mathbf{S}_e$ of the random vector $\mathbf{y}$ under H_0 and H_1 can be utilized to detect the dependence of the initial p random variables involved in $\mathbf{x}$. The main advantage of investigating $\mathbf{y}$ instead of $\mathbf{x}$ is that $\mathbf{y}$ possesses correlated structures if the components of $\mathbf{x}$ are uncorrelated. An example which studies the autoregressive conditional heteroscedastic (ARCH) type dependent structure is demonstrated in Section 4. A random vector with the ARCH type dependent structure has zero correlations and simulation results show that our test statistic based on the sample covariance matrix $\mathbf{S}$ of the original vector $\mathbf{x}$ does not work in this case.

More precisely, in this paper we consider the following linear spectral statistic

$$\int \phi(x) dF^{\mathbf{S}_e}(x) = \frac{1}{p}\sum_{j=1}^{p} \phi(\lambda_j),$$

where $\lambda_j, j = 1, 2, \ldots, p$ are the eigenvalues of the matrix $\mathbf{S}_e$ and $\phi : \mathbb{R} \to \mathbb{R}$ is an analytic function on the support of the LSD $F_c(x)$ of the matrix $\mathbf{S}_e$. Hence our test statistic is

$$LS_e = \int \phi(x) d\big(F^{\mathbf{S}_e}(x) - F_c(x)\big),$$

where $F_c(x)$ is M-P law.

The asymptotic theory for the proposed statistic LS_e under the null hypothesis is presented in the following section. It also provides a theoretical base in the high dimension situation for many test statistics in the literature proposed for the spherical test $H_0^s : \mathbf{\Sigma} = \sigma^2\mathbf{I}_p$ or the identity test $H_0^I : \mathbf{\Sigma} = \mathbf{I}_p$ since these are actually our test statistics defined by the sample covariance matrix of the original data. For example,

$$\begin{aligned} MLR &= tr\mathbf{S} - log|\mathbf{S}| - p \\ &= p\Big(\int x dF^{\mathbf{S}}(x) - \int log(x) dF^{\mathbf{S}}(x) - 1\Big), \quad \text{(Sugiura and Nagao (1968))} \end{aligned}$$

$$
\begin{aligned}
NA &= \frac{1}{p}tr(\mathbf{S}-\mathbf{I})^2 \\
&= \int (x-1)^2 dF^{\mathbf{S}}(x), \quad \text{(Nagao (1973))} \\
LW &= \frac{1}{p}tr(\mathbf{S}-\mathbf{I})^2 - \frac{p}{n}\Big[\frac{1}{p}tr\mathbf{S}\Big]^2 + \frac{p}{n} \\
&= \int (x-1)^2 dF^{\mathbf{S}}(x) - \frac{p}{n}\Big(\int x dF^{\mathbf{S}}(x)\Big)^2 + \frac{p}{n}, \quad \text{(Ledoit and Wolf (2002))}
\end{aligned}
$$

where $F^{\mathbf{S}}(x)$ is the empirical spectral distribution of the covariance matrix $\mathbf{S}$ of the investigated random vector $\mathbf{x}$. Simulation results in Section 4 show that the proposed statistic LS_e is more powerful than MLR, NA and LW in detecting nonlinear dependence, especially martingale difference sequences.

3 Main Results

In this section, the asymptotic theory for the proposed statistic LS_e is developed based on the large dimensional random matrix theory. As a matter of fact, we establish one more general result concerning the central limit theorem of linear spectral statistics for sample covariance matrices under the independent assumption.

Let $\mathbf{Z}_n = [Z_{ij}]_{p\times n}$ be a random matrix with independent complex variables. We consider the sample covariance matrix $\mathbf{B}_n = \frac{1}{n}\mathbf{Z}_n\mathbf{Z}_n^*$ under the following basic assumptions.

(**a**) For each n, $Z_{ij} = Z_{ij}^{(n)}$, $1 \le i \le p, 1 \le j \le n$ are i.i.d complex variables. And the variables satisfy the following moment conditions

$$
EZ_{11} = 0,\ E|Z_{11}|^2 = 1,\ |EZ_{11}^2| = \Phi,\ E|Z_{11}|^4 = \Psi. \tag{3.1}
$$

(**b**) Assume $p =: p(n)$ and $p/n \to c \in (0, +\infty)$.

Moreover, we denote $c_n =: p/n$ and F_{c_n} to be the MP law with parameter c_n. Set

$$
G_n(x) =: p\Big(F^{\mathbf{B}_n}(x) - F_{c_n}(x)\Big)
$$

and for any test function $f(x)$, we write

$$
L_n(f) = \int f(x) dG_n(x).
$$

Then we have the following theorem.

Theorem 1. *Assume $\mathbf{Z}_n = [Z_{ij}]_{p\times n}$ satisfy the conditions (**a**) and (**b**). Let $f_1, \dots, f_k$ be functions analytic on an open region containing the interval $[I_{(0,1)}(1-\sqrt{c})^2, (1+\sqrt{c})^2]$. Then the random vector*

$$
\Big(L_n(f_1), L_n(f_2), \dots, L_n(f_k)\Big)
$$

forms a tight sequence in n and converges weakly to a Gaussian vector $L_{f_1}, L_{f_2}, \dots, L_{f_k}$ with mean vector $\boldsymbol{\mu}$ whose components are, for any $\ell = 1, 2, \dots, k$,

$$
EL_{f_\ell} = -\frac{1}{2\pi i}\int_{\mathcal{C}} f_\ell(z)\frac{c\big(1+zs(z)\big)^3}{1-c\big(1+zs(z)\big)^2}\Big[(\Psi-\Phi^2-2) + \frac{\Phi^2}{1-\Phi^2 c\big(1+zs(z)\big)^2}\Big]dz \tag{3.2}
$$

and covariance matrix $\widetilde{\mathbf{\Sigma}}$ whose units are, for any $\ell, \nu = 1, 2, \ldots, k$,

$$\begin{aligned}
& Cov(Z_{f_\ell}, Z_{f_\nu}) \\
= \; & -\frac{1}{4\pi^2} c(\Psi - \Phi^2 - 2) \int_{\mathcal{C}_1} \int_{\mathcal{C}_2} f_\ell(z_1) f_\nu(z_2) \big(z_1 s^{'}(z_1) + s(z_1)\big)\big(z_2 s^{'}(z_2) + s(z_2)\big) dz_1 dz_2 \\
& -\frac{1}{4\pi^2} \int_{\mathcal{C}_1} \int_{\mathcal{C}_2} f_\ell(z_1) f_\nu(z_2) \frac{\Phi^2 c\big(z_1 s^{'}(z_1) + s(z_1)\big)\big(z_2 s^{'}(z_2) + s(z_2)\big)}{\big(1 - \Phi^2 a(z_1, z_2)\big)^2} dz_1 dz_2 \\
& -\frac{1}{4\phi^2} \int_{\mathcal{C}_1} \int_{\mathcal{C}_2} f_\ell(z_1) f_\nu(z_2) \frac{c\big(z_1 s^{'}(z_1) + s(z_1)\big)\big(z_2 s^{'}(z_2) + s(z_2)\big)}{\big(1 - a(z_1, z_2)\big)^2} dz_1 dz_2, \qquad (3.3)
\end{aligned}$$

where

$$a(z_1, z_2) = c\big(1 + z_1 s(z_1)\big)\big(1 + z_2 s(z_2)\big).$$

The contours in (3.2) and (3.3) are contained in the analytic region of the functions $f_1, f_2, \ldots, f_k$ and enclose the support of $F_{c_n}(x)$ for all large n. Moreover, $\mathcal{C}_1$ and $\mathcal{C}_2$ are selected to be disjoint.

Remark 1. *This theorem is a generalization of the result of Bai and Silverstein (2004) who assume $EZ_{ij}^2 = 0$ and $E(|Z_{ij}|^4) = 2$.*

Recall that $\mathbf{S}_e = \frac{1}{n}\mathbf{Y}\mathbf{Y}^*$ and the expectation of the components equals to $Ee^{itX_{jk}} = \mu(t)$ which is not zero. So our investigated matrix does not satisfy the condition (**a**). However, comparing this matrix and the corresponding normalized matrix, we can derive the limiting distributions of the linear spectral statistics for the matrix $\mathbf{S}_e$ in the following theorem.

Theorem 2. *In Theorem 1, let $Z_{jk} = e^{itX_{jk}}$ and assume conditions (**a**) and (**b**) are satisfied except that $EZ_{jk} = Ee^{itX_{jk}} = \mu(t)$. Denote $\sigma(t) := \big(var(e^{itX_{jk}})\big)^{1/2}$. Let $f_1, \ldots, f_k$ be functions analytic on an open region containing the interval $[I_{(0,1)}(1 - \sqrt{c})^2, (1 + \sqrt{c})^2]$. Then the random vector*

$$\Big(L_n(f_1), L_n(f_2), \ldots, L_n(f_k)\Big)$$

forms a tight sequence in n and converges weakly to a Gaussian vector $L_{f_1}, L_{f_2}, \ldots, L_{f_k}$ with mean vector $\boldsymbol{\mu}$ whose components are, for any $\ell = 1, 2, \ldots, k$,

$$\begin{aligned}
EZ_{f_\ell} \;=\; & -\frac{1}{2\pi i} \int_{\mathcal{C}} f_\ell(\sigma(t)z) \frac{c\big(1 + zs(z)\big)^3}{1 - c\big(1 + zs(z)\big)^2} \Big[(\Psi - \Phi^2 - 2) + \frac{\Phi^2}{1 - \Phi^2 c\big(1 + zs(z)\big)^2}\Big] dz \\
& + \frac{c}{2\pi i} \int_{\mathcal{C}} f_\ell(\sigma(t)z) \frac{-\frac{1-c}{z} + cs(z)}{z\big((1 - \frac{1-c}{z} + cs(z))^2 - c(-\frac{1-c}{z} + cs(z))^2\big)} dz \\
& - \frac{1}{2\pi i} \int_{\mathcal{C}} f_\ell(z) \frac{\int \frac{1}{(\lambda - z)^2} dF_c(\lambda)}{\big(-\frac{1-c}{z} + cs(z)\big)} dz \qquad (3.4)
\end{aligned}$$

and covariance matrix $\widetilde{\mathbf{\Sigma}}$ whose units are defined as (3.3) replacing $f_\ell(z)$ by $f_\ell(\sigma(t)z)$. Moreover, the contours in (3.4) is contained in the analytic region of the functions $f_1, f_2, \ldots, f_k$ and enclose the support of $F_{c_n}(x)$ for all large n.

The proof of Theorem 1 and Theorem 2 is relegated to the Appendix.

Remark 2. *The limit in Theorem 2 is a normal distribution which contains the unknown parameters* $\mu(t), \sigma(t), \Phi(t)$ *and* $\Psi(t)$. *These parameters have consistent estimators constructed by the sample of size n for each random variable under investigation. However, if the data are normalized at first, i.e. we consider the matrix* $\widetilde{\mathbf{S}}_e = \frac{1}{n}\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^*$ *with* $\widetilde{\mathbf{y}}_k = ((e^{itX_{1k}} - \mu(t))/\sigma(t), (e^{itX_{2k}} - \mu(t))/\sigma(t), \ldots, (e^{itX_{pk}} - \mu(t))/\sigma(t))'$ *and substitute the parameters* $\mu(t)$ *and* $\sigma(t)$ *with their consistent estimators respectively, the limiting distribution of the corresponding linear spectral statistics will be different from that with the true parameters. This is due to curse of dimensionality.*

Remark 3. *Our proposed test statistic requires that all the p tested random variables are identically distributed. It looks restrictive. But in fact, this requirement is reasonable since the provided approach is able to capture all nonlinear dependence which usually relies on high-order moments of the original variables. From another point of view, to our knowledge, almost all modern literature devoted to detecting linear dependence, i.e. spherical test or identity test for covariance matrices, require the fist four moments of all the examined variables to match the corresponding normal ones.*

4 Finite Sample Performance

4.1 Size

We now assess the finite sample performance of the proposed statistic LS_e by calculating the empirical size. The procedure is summarized as follows:

1. Generate n realizations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ from a given distribution $\mathbf{F} : \mathbb{R}^p \to \mathbb{R}$ under H_0, where $\mathbf{F}$ is the joint distribution of the investigated p-dimensional random vector $\mathbf{x} = (X_1, X_2, \ldots, X_p)$.

2. Calculate the value of the proposed statistic LS_e by the generated data in step 1, i.e. $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$.

3. Repeat step 1 and step 2 by $K = 1000$ times and we can derive K observed values for LS_e.

4. Count the number k of all the K observed values that are bigger than $z_{1-\alpha/2}$ or smaller than $z_{\alpha/2}$, where $z_{1-\alpha/2}$ and $z_{\alpha/2}$ are the $100(1-\alpha/2)\%$ and $100(\alpha/2)\%$ quantiles of the asymptotic distribution of the statistic LS_e respectively. Then the empirical size is evaluated to be $\hat{\alpha} = k/K$.

In our simulations, we consider several common distributions: (1)$\{X_i\}_{i=1}^p \sim i.i.d\ N(0,1)$; (2)$\{X_i\}_{i=1}^p \sim i.i.d\ lognormal(0,1)$; (3)$\{X_i\}_{i=1}^p \sim i.i.d\ Cauchy(0,1)$; (4)$\{X_i\}_{i=1}^p \sim i.i.d\ Burr(1,1,2)$; (5)$\{X_i\}_{i=1}^p \sim i.i.d\ Weibull(1,2)$ and (6)$\{X_i\}_{i=1}^p \sim i.i.d\ Pareto(1,1)$. These distributions are heavy-tailed distributions except $N(0,1)$.

Table 1 and Table 2 illustrate the empirical sizes for six kinds of different distributed data. The parameter t is chosen as $t = 1$. The results show the effectiveness of the proposed statistic for relatively small values of p and n. Moreover, rather rapid convergence of the empirical sizes appears as n, p increase.

As to the calculation of the quantiles $z_{\alpha/2}$ and $z_{1-\alpha/2}$ here, the procedure requires some additional work than usual. Although our asymptotic distribution is normal distribution, the asymptotic mean and variance possess relatively complicated expressions, i.e. (3.4) and (3.3).

The expressions (3.4) and (3.3) contain the moments $\sigma(t) := \big(var(e^{itX_{11}})\big)$, $\Phi = |Ee^{2itX_{11}}|$, $\Psi = E|e^{itX_{11}}|^4$ and the Stieltjes transform $s(z)$ of M-P law and its derivative $s^{'}(z)$. In our simulation, the moments of transformed data are calculated by one numerical integral function called 'quad' in MATLAB software directly since the distributions of their initial data are known which are listed above. Bai and Silverstein (2004) provides an explicit expression for the Stieltjes transform as follows

$$\underline{s}(z) = \frac{-(z+1-c) + \sqrt{(z-1-c)^2 - 4c}}{2z}$$

and $s(z)$ is related to $\underline{s}(z)$ by

$$s(z) = \frac{1}{-z - z\underline{s}(z)}.$$

Then obviously the explicit expression of its derivative $s^{'}(z)$ can be derived. For the several contour integrals appeared in the asymptotic mean and variance in (3.4) and (3.3), we utilize polar coordinates and variable substitutions $z = re^{i\theta}$ with r and θ changing from 0 to 10 and 0 to 2π respectively. With contour integrals replaced by double integrals, we use the numerical integral function named 'dblquad' in MATLAB to calculate them.

Remark 4. *In simulations, the first four moments of transformed data can be calculated directly since the distributions of the initial data are available. However, we can also use the sample moments to substitute the corresponding population moments since they are consistent estimations.*

To examine the effect of the choice of the parameter t, we consider t in $[0.1, 2]$. Figure 1 reports the empirical sizes for various t in the scenario of $(n, p) = (100, 60)$. It shows that the empirical size is not very sensitive to the value of the parameter t except when t is near to zero. It can be seen that the empirical size equals to zero if t is 0.1 while a big jump at $t = 0.2$. This phenomenon can be explained by the fact that small t which is near zero would cause all the components of the transformed data close to 1 so that the randomness is lost. From these graphs, we suggest to choose t from the range $[0.3, 1.5]$ in practice.

4.2 Power

To investigate power, we consider the following three data generating processes (DGPs).

1. DGP1: Auto-Regressive Conditional Heteroskedasticity of first order (ARCH(1)) model:

$$X_t = \varepsilon_t h_t^{1/2},\ h_t = 1 + 0.8X_{t-1}^2, t = 1, 2, \ldots, p;$$

2. DGP2: Moving-Average of first order$\big($MA(1)$\big)$ model:

$$X_t = 0.8\varepsilon_{t-1} + \varepsilon_t, t = 1, 2, \ldots, p;$$

3. DGP3: Auto-Regressive of first order$\big($AR(1)$\big)$ Model:

$$X_t = 0.2X_{t-1} + \varepsilon_t, t = 1, 2, \ldots, p;$$

where we consider two innovation processes: (1) $\{X_0, \varepsilon_t\} \sim i.i.d\ N(0,1)$ and $\{X_0, \varepsilon_t\} \sim i.i.d\ lognormal(0,1)$.

For each DGP, we generate $n + 100$ observations and then discard the first 100 in order to mitigate the impact of the initial values. Implement the three steps used in calculating empirical sizes and derive the empirical power as $\beta = k/K$. The powers are listed in Table 3. It is well known that ARCH(1) model is a martingale difference sequence and the components of this series are uncorrelated but dependent. From the table, we can see that while LS_e is effective for this model, the statistic LS fails to capture their dependence due to rather low powers comparable to the significant level. This phenomenon maybe reflects that the limiting spectral distribution of the sample covariance matrix for a martingale difference sequence is also the M-P law. However, this is not provided with any theoretical proof from random matrix theory yet. By converting dependent relationships into correlated structures, our proposed statistic is powerful in detecting dependence in practice. For the other two models, both LS_e and LS work well since their dependence is measured by correlation.

It remains to see if the power is robust under different choices of the parameters t. We consider t in the range $[0.1, 1.5]$. From Figure 2, we can see that powers do not change greatly with various values of t.

5 Empirical Applications

We now apply the proposed methods to the daily returns of the 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 j at day τ is denoted as $S_{j\tau}$. These data are derived from Wharton Research Data Services (WRDS). We use the logarithmic difference $X_{j\tau} = ln(S_{j\tau}/S_{j,\tau-1})$. Then $N = 251$ daily returns are available for each stock. Note that although we have $N = 251$ samples available for each stock here, we usually only utilize the first n samples to do the test with $n \leq N$. The value of n depends on the value of p in order to making them at the same order.

The interest here is to test whether the daily returns for the investigated p stocks are dependent. Here we investigate three groups of companies, i.e. $p = 10, 40, 60$ stocks respectively from S&P500. We need that every stock has the same distribution. In order to derive identically distributed data from the original data, we assume that each stock return series satisfies the general normal distribution, i.e. the transformation of the data follows a standard normal distribution,

$$\hat{X}_{jt} := \Big(\frac{X_{jt} - a_j}{b_j}\Big)^{\beta_j} \sim Normal(0,1), \tag{5.1}$$

where a_j, b_j, β_j are unknown parameters. This distribution possesses high peak and heavy tails compared with the normal distribution. It is a typical property of the financial data (Rama (2001)). Figure 3 illustrates the smoothed empirical densities of the transformed data for all the selected 96 stocks under investigation. From these graphs, we can see that the model (5.1) is fitted well for stock returns.

To utilize the proposed statistic, we transform $\hat{X}_{jt}$ to $\widetilde{X}_{j\tau} = e^{it\hat{X}_{j\tau}}$. For each stock, its daily return series $\{\widetilde{X}_{j1}, \widetilde{X}_{j2}, \ldots, \widetilde{X}_{jn}\}$ can be assumed to be independent. In the financial literature, this assumption is always satisfied by stock returns when they are calculated by a large time scale, for example, one day.

With the processed data at hand, the observed LS_e are calculated. We randomly choose p companies from the total available 96 companies and calculate the proposed statistic values. Repeat this experiment with $K = 5$ times and the number of $K = 5$ statistic values are derived. They are listed in Table 4. From this table, we can see that the more companies are included, the larger the statistic values will be. In the case of $(p, n) = (10, 20)$, all the five statistic values are included in the interval with critical values as two end points. We should accept the null hypothesis that the randomly chosen 10 stocks are independent. Under the other two cases, obviously the statistic values are outside the interval bounded by the critical values and we can conclude that randomly chosen 40 or 60 stocks are dependent.

As to the impact of the choice of the parameter t in LS_e on the statistic values, we randomly choose 10 companies with sample size of $n = 20$ and plot the graphs of the p-values against different t values in the range of $0.1 - 1.5$. From Figure 5, it can be seen that the p-values are not very sensitive to the choice of the parameter t. Moreover, most of the selected t values tell us rather high p-values and thus we should accept the null hypothesis. Note that the p-value corresponding to $t = 0.1$ is 0. This value should not be adopted in practice since the empirical size corresponding to $t = 0.1$ is not well.

6 Conclusion

By replacing the original data by its 'characteristic' type transformed data, we utilize the linear spectral statistic based on the sample covariance matrix of the transformed data to detect the dependence of the original data in a high dimensional scenario. These generalized spectral can capture more types of dependent structures, such as ARCH(1) model, which is one typical example on behalf of uncorrelated but dependent relationships. Moreover, this statistic requires no moment condition and is well defined for both discrete and continuous random variables. This advantage makes it effective for heavy-tailed data. Another key advantage of this methodology is that it requires no prior information of any moments or the distribution of the original data. The asymptotic theory for the proposed statistic under the null hypothesis is established. As one contribution of this paper in random matrix theory, the asymptotic theorem generalizes the central limit theorem for linear spectral statistics in Bai and Silverstein (2004) by eliminating the condition that the second moment of complex random variables is zero. The extensive simulation and an empirical application to daily stock returns in S&P500 highlight this approach.

References

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

Adler, R., Feldman, R. and Taqqu, M.(1998). A practical guide to heavy tails: statistical techniques and applications. *Birkhäuser* **1st edition**.

Bai, Z. D. and Silverstein, J. W.(2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32(1A)**, 553-605.

Bai, Z. D. and Silverstein, J. W.(2006). Spectral analysis of large dimensional random matrices. Mathematics Monograph Series 2, Science Press, Beijing.

Bai, Z. D., Jiang, D. D., Yao, J. F. and Zheng, S. R.(2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37(6B)**, 3822-3840.

Bai, Z.D., Miao, B.Q., Pan, G.M.(2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35(4)**, 1532-1572.

Bai, Z.D., Yao, J.F.(2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11(6)**, 1059-1092.

Bai, Z. D. and Zhou, W.(2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica.* **18**, 425-442.

Bao, Z.G. (2018). Tracy-Widom limit for Spearman's rho. (Preprint)

Bao, Z.G. (2019). Tracy-Widom limit for Kendall's tau. *The Annals of Statistics*, **47**(6), pp.3504-3532.

Bao, Z.G., Pan, G.M., and Zhou, W. (2012). Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. *Electronic Journal of Probability*, **17**1-32.

Bao, Z. G., Lin, L. C., Pan, G.M., and Zhou, W. (2015). Spectral statistics of large dimensional Spearman?s rank correlation matrix and its application. *The Annals of Statistics*, **43**(6), 2588-2623.

Billingsley, P.(1999). Convergence of probability measures, *Wiley-Interscience Publication*, **second edition**.

Cai, T. T. and Jiang, T. F.(2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39(3)**, 1496-1525.

Fan, J. and Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. In *International Congress of Mathematicians* **III 595-622. Eur. Math. Soc., Zrich**.

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.

Johnstone, I. M.(2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, 295-327.

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

Liu, W. D., Lin, Z. Y. and Shao, Q. M.(2008). The asymptotic distribution and BerryEsseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18(6)**, 2337-2366.

Lukacs, E.(1970). Characteristic functions. *London: Charles Griffin* **2nd edition**.

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

Nagao, H.(1973). On some test criteria for covariance matrix. *Ann. Statist.* **1**, 700-709.

Pan, G.M.(2011). Comparison between two types of large sample covariance matrices. *Annales de l′Institut Henri Poincaré*, **http://www.ntu.edu.sg/home/gmpan/publications.html**.

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

Pan, G.M. and Zhou, W.(2011). Central limit theorem for Hotelling's T^2 statistic under large dimension. *Ann. Appl. Probab.* **21(5)**, 1860–1910.

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

Schott, J. R.(2005). Testing for complete independence in high dimensions. *Biometrika* **92**, 951-956.

Sugiura, N. and Hisao Nagao(1968). Unbiasedness of some test criteria for the equality of one or two covariance matrices. *Ann. Math. Statist.***39**, 1686-1692.

Tong, H.(1990). Nonlinear time series: A Dynamic System Approach. *Oxford, U.K.: Clarendon Press.*

Zhou, W.(2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359**, 5345-5363.

7 Appendix

In this Appendix, we will sketch the proofs of Theorem 1 and Theorem 2. The details of the proofs will be presented in the supplementary material.

7.1 Outline of proof of Theorem 1

Note that Theorem 1 is a generalization of Theorem 1.1 of Bai and Silverstein (2004) (under the setting of $T_n = I$) by removing the condition $\mathbb{E}(Z_{11}^2) = 0$ and $\mathbb{E}|Z_{11}|^4 = 2$. Since the condition $\mathbb{E}|Z_{11}|^4 = 2$ has been removed in Pan and Zhou (2008), it suffices to remove the condition $\mathbb{E}(Z_{11}^2) = 0$. Therefore, all the discussions that do not need $\mathbb{E}\{Z_{11}^2\} = 0$ can be borrowed from Bai and Silverstein (2004) and Pan and Zhou (2008) directly. The basic idea raised in Bai and Silverstein (2004) is to use the Cauchy integral formula to write

$$L_n(f) = \int f(x)dG_n(x) = -\frac{1}{2\pi i}\int_{\mathcal{C}} f(z)s_{G_n}(z)dz$$

with an appropriately chosen contour enclosing the support of $G_n(x)$. Then Bai and Silverstein (2004) showed that to study the weak convergence of $(L_n(f_1), \cdots, L_n(f_k))$, it suffices to establish the weak convergence of the process $\{\widehat{M}_n(z), \mathcal{C}\}$. We refer to (1.11) of Bai and Silverstein (2004) for the definition of $\widehat{M}_n(z)$. In fact, $\widehat{M}_n(z)$ is a modification of $M_n(z) = ps_{G_n}(z)$. It means that we can transfer the problem into the study of the weak convergence of $M_n(z)$. For ease of discussion, in the supplementary material, we adopt a slightly modified strategy which has been used in Bai and Yao (2005). The strategy is to derive the weak convergence for the process $\{M_n(z), \mathbb{C}_0\}$ at first, and then extend the issue to the whole contour $\mathcal{C}$. Here $\mathbb{C}_0 = \{z = u + iv : |v| \geq v_0\}$ for some $v_0 > 0$. Note that by the definition in Bai and Silverstein (2004), on the region $\mathbb{C}_0$, $\widehat{M}_n(z) = M_n(z)$. Though the proof in Bai and Yao (2005) is given for Wigner matrices, it can be easily transferred to the sample covariance matrices. Moreover, since the extension from $\mathbb{C}_0$ to the whole $\mathcal{C}$ does not need the assumption $\mathbb{E}(Z_{11}^2) = 0$, we omit this part in the supplementary material. To study the weak convergence of $\{M_n(z), \mathbb{C}_0\}$, we split $M_n(z)$ into the random part

$$M_n^1(z) = p(s_n(z) - \mathbb{E}s_n(z))$$

and the non-random part

$$M_n^2(z) = p(\mathbb{E}s_n(z) - s_n^0(z)).$$

Here $s_n(z)$ is the Stieltjes transform of $\mathbf{B}_n$ and $s_n^0(z)$ is the Stiltjes transform of the MP law with parameter $c_n = p/n$. Therefore it suffices to provide the tightness and the finite dimensional convergence of $M_n^1(z)$ and the limit of $M_n^2(z)$ on $\mathbb{C}_0$. Actually, the only differences between our results from those in Bai and Silverstein (2004) lie on the covariance function of $M_n^1(z)$ and the limit of $M_n^2(z)$. Therefore, in the supplementary material, we provide the detailed calculations for them. The main differences come from the identity (1.15) of Bai and Silverstein (2004). Without the assumption $\mathbb{E}(Z_{11})^2 = 0$, one needs to evaluate the expectations or conditional expectations of the quantities in the form of $tr\mathbf{AB}'$. More precisely, when we evaluate the covariance function of $M_n^1(z)$, it is

$$\frac{1}{n} tr\mathbb{E}_j(\mathbf{D}_j^{-1}(z_1))\mathbb{E}_j(\mathbf{D}_j^{-1}(z_2))'.$$

For the limit of $M_n^2(z)$, it is

$$\frac{1}{n}\mathbb{E}tr\mathbf{D}_j^{-1}(z)(\mathbf{D}_j^{-1}(z))'.$$

See the supplementary material for the definition of $\mathbf{D}_j(z)$. Section 1 of the supplementary material is mainly devoted to the estimations of these two quantities. We claim all the other parts of the proof of Theorem 1 do not depend on the assumption $\mathbb{E}(Z_{11}^2) = 0$, and thus they are exactly the same as the original proof in Bai and Silverstein (2004) and Pan and Zhou (2008). So we do not present the details in the supplementary material and just sketch some main steps if necessary.

7.2 Outline of proof of Theorem 2

For all $j = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, n$, denote $\mathbf{z}_k = (Z_{1k}, Z_{2k}, \ldots, Z_{pk})'$, where $Z_{jk} = \big(e^{itX_{jk}} - \mu(t)\big)/\sigma(t)$ with $\mu(t) = Ee^{itX_{jk}}$ and $\sigma = \big(Var(e^{itX_{jk}})\big)^{1/2}$. Then the transformed data can be written as

$$\mathbf{y}_k = \sigma(t)\mathbf{z}_k + \boldsymbol{\mu}(t), \tag{7.1}$$

where $\boldsymbol{\mu}(t) = (\mu_1(t), \mu_2(t), \ldots, \mu_p(t))$ with $\mu_j(t) = \mu(t)$ for all $j = 1, 2, \ldots, p$.

Write

$$\bar{\mathbf{y}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{y}_j,\ \ \bar{\mathbf{z}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{z}_j,$$

and

$$\mathbf{D}_e = \frac{1}{n}\sum_{j=1}^{n}\mathbf{z}_j\mathbf{z}_j^*,\quad \mathcal{S}_e = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{y}_j-\bar{\mathbf{y}})(\mathbf{y}_j-\bar{\mathbf{y}})^*,\quad \mathcal{D}_e = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{z}_j-\bar{\mathbf{z}})(\mathbf{z}_j-\bar{\mathbf{z}})^*.$$

Note that $\mathcal{S}_e = \mathcal{D}_e$. The investigated sample covariance matrix $\mathbf{S}_e$ can be then expressed as

$$\mathbf{S}_e = \mathcal{S}_e + \bar{\mathbf{y}}\bar{\mathbf{y}}^* = \mathcal{D}_e + \bar{\mathbf{y}}\bar{\mathbf{y}}^*. \tag{7.2}$$

By Cauchy's formula, with probability one, for n large,

$$\begin{aligned} p\int f(x)d[F^{\mathbf{S}_e}(x) - F_{c_n}(x)] &= \frac{p}{2\pi i}\int\oint_{\gamma}\frac{f(z)}{z-x}dzd[F^{\mathbf{S}_e}(x) - F_{c_n}(x)] \\ &= \frac{p}{2\pi i}\oint_{\gamma} f(z)dz\int\frac{1}{z-x}d[F^{\mathbf{S}_e}(x) - F_{c_n}(x)] \\ &= -\frac{1}{2\pi i}\oint_{\gamma} f(z)(tr(\mathbf{S}_e - z\mathbf{I}_p)^{-1} - ps_n^0(z))dz, \end{aligned} \tag{7.3}$$

where $s_n^0(z)$ is obtained from $s_c(z)$ with c replaced by c_n. The contour γ is specified as follows: Let $v_0 > 0$ be arbitrary and set $\gamma_\mu = \{\mu + iv_0, \mu \in [\mu_\ell, \mu_r]\}$, where $\mu_r > (1+\sqrt{c})^2$ and $0 < \mu_\ell < I_{(0,1)}(c)(1-\sqrt{c})^2$ or μ_ℓ is any negative number if $c \geq 1$. Then define

$$\gamma^+ = \{\mu_\ell + iv : v \in [0, v_0]\} \cup \gamma_\mu \cup \{\mu_r + iv : v \in [0, v_0]\}$$

and let γ^- be the symmetric part of γ^+ about the real axis. Then set $\gamma = \gamma^+ \cup \gamma^-$.

Set

$$\begin{aligned} \mathcal{S}_e^{-1}(z) = (\mathcal{S}_e - z\mathbf{I}_p)^{-1},\ \mathbf{S}_e^{-1}(z) = (\mathbf{S}_e - z\mathbf{I}_p)^{-1}, \\ \mathcal{D}_e^{-1}(z) = (\mathcal{D}_e - z\mathbf{I}_p)^{-1},\ \mathbf{D}_e^{-1}(z) = (\mathbf{D}_e - z\mathbf{I}_p)^{-1}. \end{aligned}$$

Then we have

$$tr\mathbf{S}_e^{-1}(z) - ps_n^0(z) = \left(tr\mathcal{D}_e^{-1}(z) - ps_n^0(z)\right) - \frac{\bar{\mathbf{v}}^*\mathcal{D}_e^{-2}(z)\bar{\mathbf{v}}}{1+\bar{\mathbf{v}}^*\mathcal{D}_e^{-1}(z)\bar{\mathbf{v}}}, \tag{7.4}$$

The limiting distribution of $\left(tr\mathcal{D}_e^{-1}(z) - ps_n^0(z)\right)$ has been provided in Pan (2011). The second term $\frac{\bar{\mathbf{v}}^*\mathcal{D}_e^{-2}(z)\bar{\mathbf{v}}}{1+\bar{\mathbf{v}}^*\mathcal{D}_e^{-1}(z)\bar{\mathbf{v}}}$ converges to a constant in probability, whose detailed calculation is included in section 2 of supplementary material. Combining (7.3) with (7.4), we can derive the result of Theorem 2.

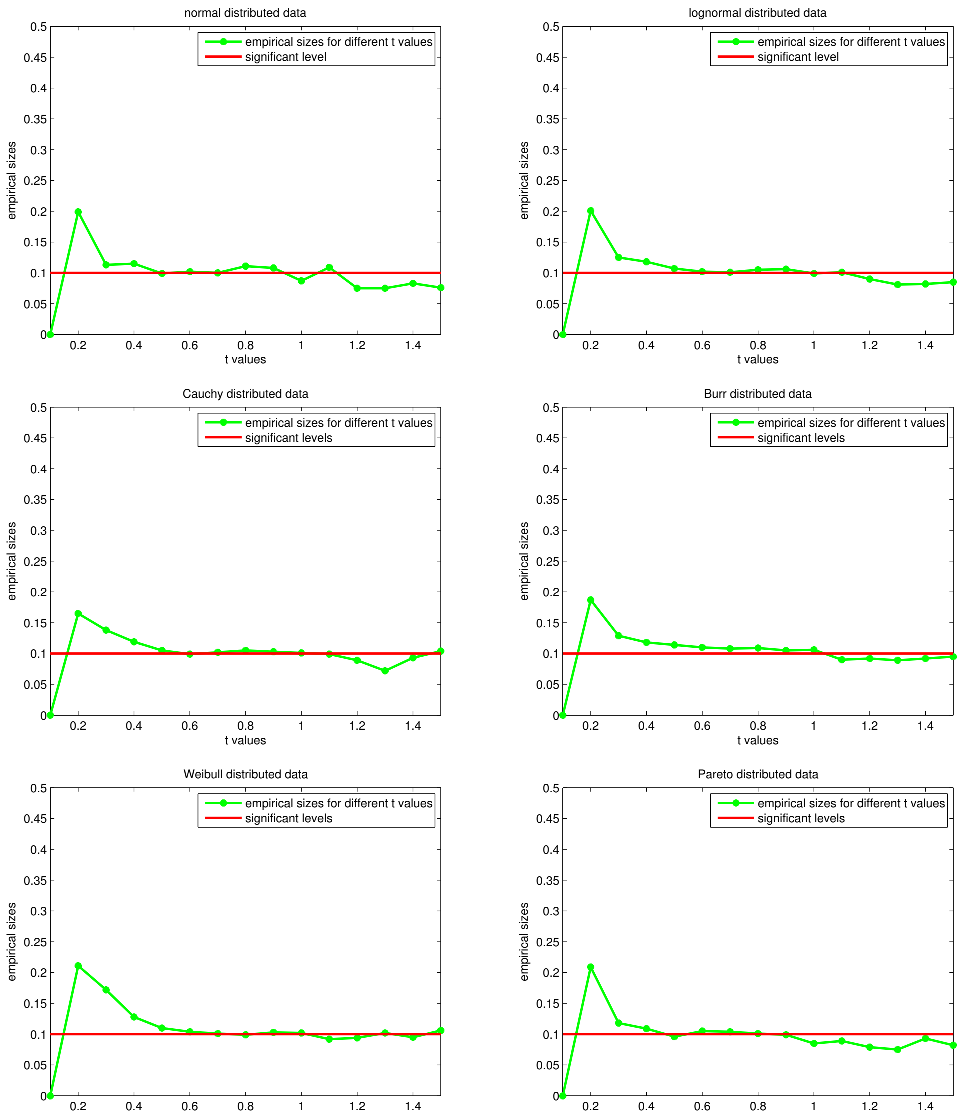
Table 1: Empirical sizes for $N(0,1)$, $lognormal(0,1)$ and $Cauchy(0,1)$ distributed data respectively under several significant levels and $t = 1$

(p,n)	Empirical sizes for the following significant levels				
	15%	10%	7.5%	5%	2.5%
Normal(0,1)					
(10,20)	0.139	0.089	0.069	0.041	0.019
(30,40)	0.142	0.090	0.071	0.042	0.022
(60,100)	0.145	0.096	0.079	0.052	0.026
(80,200)	0.146	0.104	0.076	0.053	0.024
(200,300)	0.148	0.101	0.078	0.051	0.026
(20,10)	0.140	0.087	0.070	0.042	0.020
(40,30)	0.142	0.090	0.072	0.043	0.021
(100,60)	0.146	0.092	0.078	0.051	0.023
(200,80)	0.148	0.098	0.077	0.052	0.021
(300,200)	0.149	0.102	0.076	0.051	0.024
lognormal(0,1)					
(10,20)	0.135	0.087	0.068	0.038	0.017
(30,40)	0.139	0.091	0.069	0.041	0.020
(60,100)	0.142	0.094	0.072	0.043	0.022
(80,200)	0.147	0.096	0.074	0.046	0.024
(200,300)	0.151	0.099	0.076	0.048	0.023
(20,10)	0.137	0.085	0.070	0.040	0.018
(40,30)	0.141	0.089	0.071	0.040	0.021
(100,60)	0.143	0.093	0.075	0.044	0.024
(200,80)	0.149	0.101	0.076	0.049	0.026
(300,200)	0.152	0.101	0.074	0.052	0.025
Cauchy(0,1)					
(10,20)	0.132	0.085	0.069	0.040	0.019
(30,40)	0.138	0.088	0.071	0.041	0.018
(60,100)	0.143	0.093	0.074	0.046	0.021
(80,200)	0.146	0.097	0.074	0.047	0.024
(200,300)	0.149	0.101	0.077	0.049	0.027
(20,10)	0.131	0.082	0.070	0.039	0.019
(40,30)	0.135	0.084	0.070	0.040	0.021
(100,60)	0.142	0.095	0.073	0.048	0.023
(200,80)	0.148	0.100	0.076	0.051	0.025
(300,200)	0.151	0.102	0.076	0.052	0.024

Table 2: Empirical sizes for $Burr(1,1,2)$, $Weibull(1,2)$ and $Pareto(1,1)$ distributed data respectively under several significant levels and $t = 1$

(p, n)	Empirical sizes for the following significant levels				
	15%	10%	7.5%	5%	2.5%
Burr(1,1,2)					
(10,20)	0.140	0.090	0.071	0.039	0.021
(30,40)	0.142	0.090	0.070	0.041	0.019
(60,100)	0.148	0.096	0.073	0.042	0.023
(80,200)	0.147	0.100	0.071	0.046	0.023
(200,300)	0.150	0.099	0.073	0.051	0.022
(20,10)	0.139	0.088	0.071	0.040	0.019
(40,30)	0.141	0.092	0.072	0.040	0.022
(100,60)	0.144	0.095	0.072	0.044	0.024
(200,80)	0.142	0.101	0.074	0.044	0.025
(300,200)	0.146	0.098	0.076	0.048	0.026
Weibull(1,2)					
(10,20)	0.133	0.082	0.068	0.035	0.018
(30,40)	0.138	0.089	0.070	0.039	0.019
(60,100)	0.142	0.093	0.072	0.043	0.020
(80,200)	0.145	0.095	0.075	0.047	0.024
(200,300)	0.149	0.098	0.076	0.051	0.026
(20,10)	0.132	0.084	0.070	0.037	0.017
(40,30)	0.140	0.088	0.072	0.040	0.020
(100,60)	0.144	0.096	0.076	0.043	0.022
(200,80)	0.148	0.095	0.074	0.048	0.024
(300,200)	0.150	0.094	0.074	0.047	0.024
Pareto(1,1)					
(10,20)	0.131	0.084	0.065	0.037	0.016
(30,40)	0.137	0.089	0.069	0.040	0.018
(60,100)	0.142	0.090	0.071	0.042	0.022
(80,200)	0.147	0.095	0.074	0.046	0.024
(200,300)	0.149	0.098	0.075	0.049	0.023
(20,10)	0.134	0.085	0.069	0.039	0.017
(40,30)	0.139	0.087	0.071	0.040	0.020
(100,60)	0.142	0.094	0.072	0.042	0.023
(200,80)	0.146	0.095	0.072	0.046	0.022
(300,200)	0.148	0.097	0.076	0.049	0.023

Figure 1: Graphs of empirical sizes vs t values



*The empirical sizes are calculated under the scenario of $n = 100, p = 60$.

Table 3: Empirical powers of the proposed statistic LS_e compared with the statistic LS under some significant levels

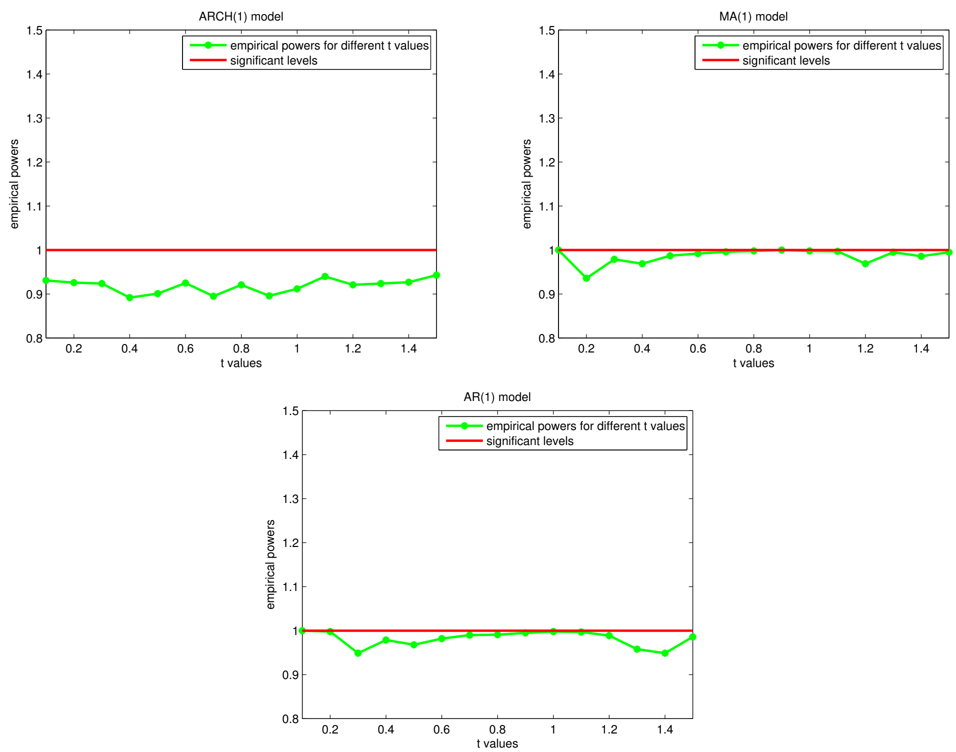
Empirical powers under significant levels				
	10%	5%	10%	5%
(p,n)	LS_e		LS	
ARCH(1,1)				
(10,20)	0.884	0.805	0.122	0.076
(30,40)	0.901	0.831	0.103	0.066
(60,100)	0.912	0.840	0.099	0.052
(80,200)	0.958	0.872	0.100	0.060
(200,300)	0.981	0.882	0.104	0.059
(20,10)	0.873	0.799	0.131	0.089
(40,30)	0.910	0.829	0.109	0.068
(100,60)	0.920	0.830	0.106	0.059
(200,80)	0.951	0.852	0.100	0.052
(300,200)	0.990	0.901	0.099	0.058
MA(1)				
(10,20)	0.944	0.932	0.927	0.909
(30,40)	0.958	0.940	0.952	0.921
(60,100)	0.994	0.978	0.991	0.948
(80,200)	1.000	0.995	1.000	0.990
(200,300)	1.000	0.994	1.000	0.992
(20,10)	0.954	0.931	0.959	0.929
(40,30)	0.960	0.935	0.962	0.938
(100,60)	0.999	0.983	0.997	0.969
(200,80)	1.000	1.000	1.000	0.999
(300,200)	1.000	1.000	1.000	1.000
AR(1)				
(10,20)	0.960	0.948	0.969	0.949
(30,40)	0.988	0.980	0.990	0.985
(60,100)	0.998	0.990	1.000	1.000
(80,200)	1.000	1.000	1.000	1.000
(200,300)	1.000	1.000	1.000	1.000
(20,10)	0.972	0.960	0.973	0.969
(40,30)	0.990	0.985	0.995	0.991
(100,60)	0.996	0.990	0.999	0.995
(200,80)	1.000	1.000	1.000	1.000
(300,200)	1.000	1.000	1.000	1.000

Table 4: Statistic values under various scenarios for randomly chosen by 5 times

(p,n)	10%critical values	1	2	3	4	5
(10,20)	[-31.5746, 27.4875]	-15.3490	3.6996	-19.9383	-1.7622	-10.4121
(30,50)	[-52.0903, 45.2941]	7803.5	1064.0	9275.2	9727.7	9924.2
(60,80)	[-64.6601, 56.2072]	93718	106970	132900	103340	91292

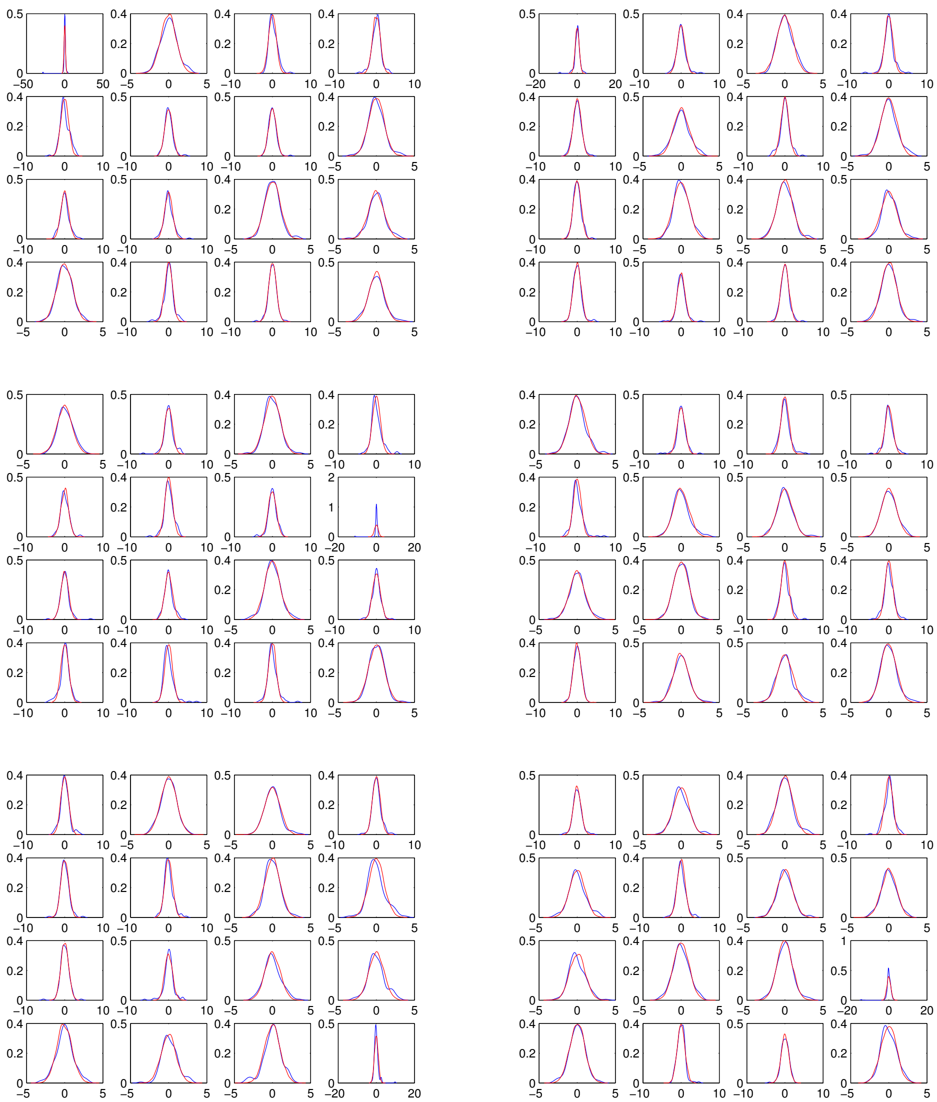
*The critical values correspond to those of the limit normal distributions of the statistic LS_e under the three cases of $(p,n) = (10,20), (30,50), (60,80)$ respectively.

Figure 2: Graphs of empirical powers vs t values



*The empirical powers are calculated under the scenario of $n = 100, p = 60$.

Figure 3: 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.

Figure 4: Graphs of autocorrelation function for randomly chosen 6 companies from all the 96 companies

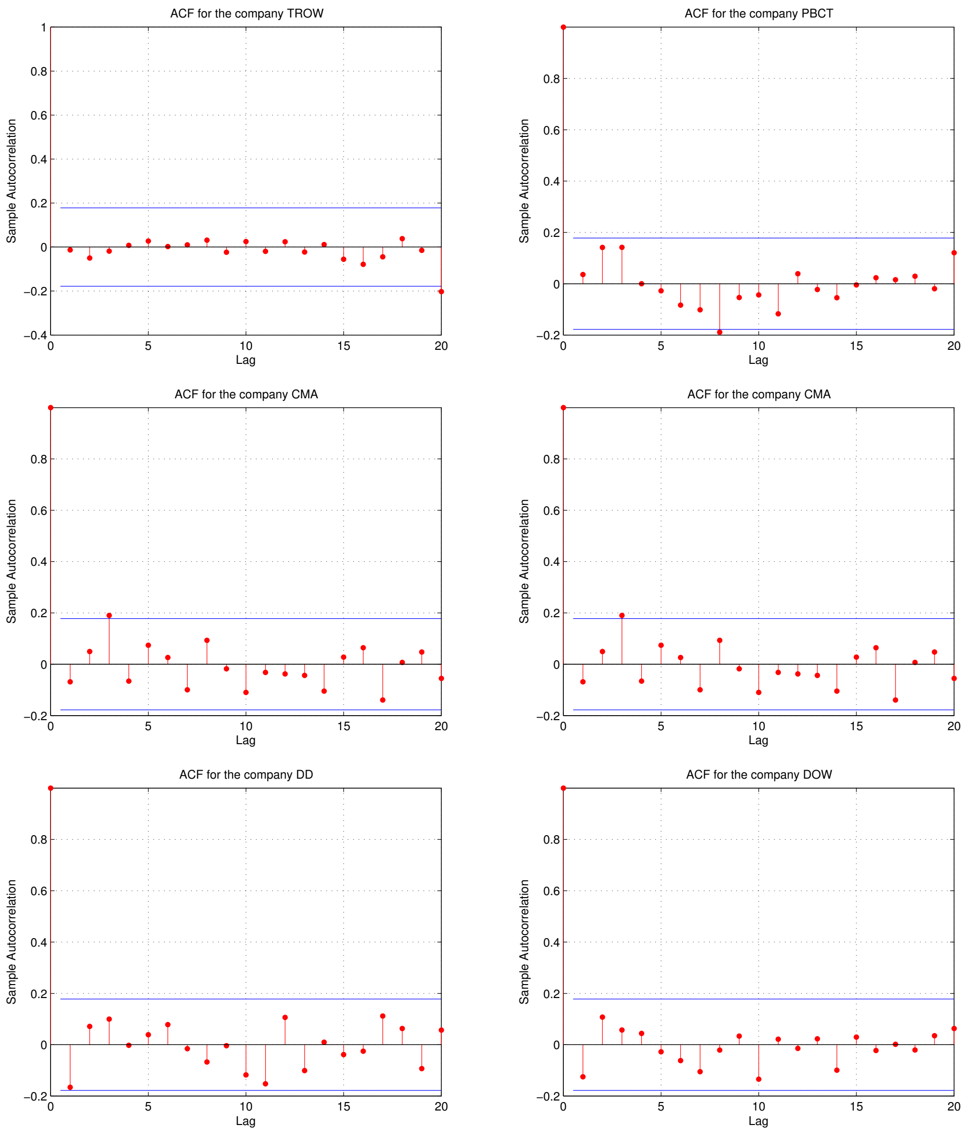
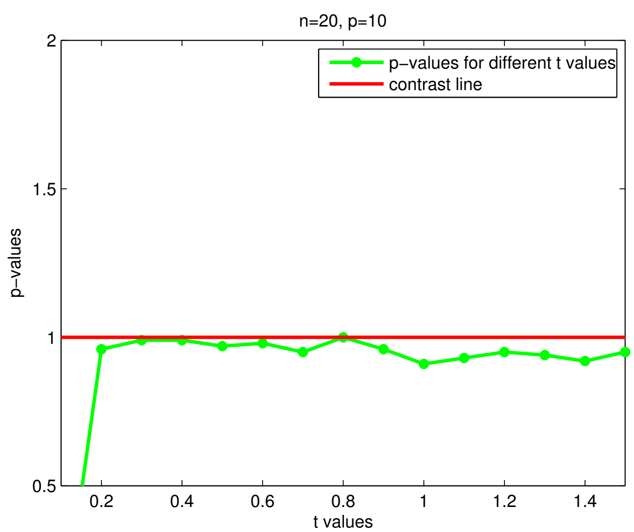


Figure 5


n=20, p=10
p−values for different t values
contrast line
p−values
t values
2
1.5
1
0.5
0.2
0.4
0.6
0.8
1
1.2
1.4