Statistica Sinica Preprint No: SS-2022-0354	
Title	Asymptotic Independence of the Sum and Maximum of Dependent Random Variables with Applications to High-dimensional Tests
Manuscript ID	SS-2022-0354
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0354
Complete List of Authors	Long Feng, Tiefeng Jiang, Xiaoyun Li and Binghui Liu
Corresponding Authors	Xiaoyun Li
E-mails	xl374@scarletmail.rutgers.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica

ASYMPTOTIC INDEPENDENCE OF THE SUM AND MAXIMUM OF DEPENDENT RANDOM VARIABLES WITH APPLICATIONS TO HIGH-DIMENSIONAL TESTS

Long Feng[1], Tiefeng Jiang[2], Xiaoyun Li[3], Binghui Liu[4]

[1] *Nankai University,* [2] *University of Minnesota,* [3] *LinkedIn Corporation and* [4] *Northeast Normal University*

Abstract: For a set of dependent random variables, without stationary or the strong mixing assumptions, we derive the asymptotic independence between their sums and maxima. Then we apply this result to high-dimensional testing problems, where we combine the sum-type and max-type tests and propose a novel test procedure for the one-sample mean test, the two-sample mean test and the regression coefficient test in high-dimensional setting. Based on the asymptotic independence between sums and maxima, the asymptotic distributions of test statistics are established. Simulation studies show that our proposed tests have good performance regardless of data being sparse or not. Examples on real data are also presented to demonstrate the advantages of our proposed methods.

Key words and phrases: Asymptotic normality; Asymptotic independence; Extreme-value distribution; High-dimensional tests; Large p and small n

1. Introduction

Statistical independence is a very simple structure and is convenient in statistical inference and applications. In this paper, we study the asymptotic independence between two common statistics: the extreme-value statistic $M_p = \max_{1\leqslant i\leqslant p} X_i$ and the sum $S_p = \sum_{i=1}^p X_i$, where $\{X_i\}_{i=1}^p$ is a sequence of dependent random variables. This theoretical results will be applied to three high-dimensional testing problems with numerical examples.

1.1 Independence Between Sum and Maximum

In the past few decades, great efforts have been devoted in understanding the asymptotic joint distribution of M_p and S_p. In an early research, Chow and Teugels (1978) established the asymptotic independence between M_p and S_p for independent and identically distributed random variables. To overcome the limitation of the required assumptions, Anderson and Turkman (1991), Anderson and Turkman (1993), Anderson and Turkman (1995) and Hsing (1995) generalized the asymptotic result to the case that $\{X_i\}_{i=1}^p$ is strong mixing; for the concept "strong mixing" and its properties, see, for example, the survey paper Bradley (2005) and the literature therein. In particular, Hsing (1995) showed that for a stationary sequence, strong mixing property and asymptotic normality of S_p are basically enough to guarantee the asymptotic independence of the sum and maximum. However, it is shown in Davis and Hsing (1995) that in the case of infinite variance, M_p and S_p are not asymptotically independent because the asymptotic behavior of S_p is dominated by that of the extreme order statistic. In addition, Ho and Hsing (1996), Ho and McCormick (1999), McCormick and Qi (2000) and Peng and Nadarajah (2003) considered the joint limit distribution of the maximum and sum of stationary Gaussian sequence $\{X_i\}_{i=1}^p$ in which $E(X_i) = 0$, $\mathrm{Var}(X_i) = 1$ and $r(p) = E(X_i X_{i+p})$. Under different conditions on $r(p)$, the joint limiting distributions of maxima and sums are different. Specifically, Ho and Hsing (1996) showed that M_p and S_p are asymptotically independent as long as $\lim_{p\to\infty} r(p)\log p = 0$; the two statistics are not independent provided $\lim_{p\to\infty} r(p)\log p = \rho \in (0,\infty)$. For the rest situations, by assuming $\lim_{p\to\infty} \frac{\log p}{p} \sum_{i=1}^p |r(i) - r(p)| = 0$, Ho and McCormick (1999) and McCormick and Qi (2000) obtained the asymptotic independence of $M_p - (S_p/n)$ and S_p.

All these results are based on the stationary assumption that the covariance structure among $\{X_i\}_{i=1}^p$ has the property that $E(X_i X_{i+h}) = E(X_1 X_{1+h})$ for each integer h and $i = 1, \cdots, p-h$. This is a common assumption in research, however, it is not easy to be checked. Even though it can be verified by hypothesis testing, the stationary property still may not hold up to certain statistical errors. In fact, in many scenarios this assumption is not true. For example, for stock data of US S&P 500 index in which stock returns are considered as variables, if stocks are ordered alphabetically by names, the two stocks, such as AAPL and MSFT, may have both far distance and strong correlation, which does not satisfy the stationary assumption.

In this work, we study the asymptotic independence between $\tilde{S}_p = \sum_{i=1}^p Z_i^2$ and $\tilde{M}_p = \max_{1\leqslant i\leqslant p} Z_i^2$ without stationary assumption. In our case each Z_i is marginally $N(0,1)$ and the covariance matrix of Z_i's, denoted by $\mathbf{\Sigma}_p = (\sigma_{ij})_{1\leqslant i,j\leqslant p}$, satisfies certain conditions. Specifically, we first establish the asymptotically normality of $\tilde{S}_p$ if $[\mathrm{tr}(\mathbf{\Sigma}_p^{2+\delta})]^2 \cdot [\mathrm{tr}(\mathbf{\Sigma}_p^2)]^{-2-\delta} \to 0$ for some $\delta > 0$. Then, we show that the limit distribution of the maximum $\tilde{M}_p - 2\log p + \log\log p$ is a Gumbel distribution under conditions on the covariance matrix $\mathbf{\Sigma}_p$. Finally, we prove the asymptotic independence between $\tilde{S}_p$ and $\tilde{M}_p$ under the conditions $\max_{1\leqslant i<j\leqslant p} |\sigma_{ij}| \leqslant \varrho$ and $\max_{1\leqslant i\leqslant p} \sum_{j=1}^p \sigma_{ij}^2 \leqslant (\log p)^C$, together with two additional conditions on the maximum and minimum eigenvalues of $\mathbf{\Sigma}_p$. These theoretical results are novel and essentially different from the existing ones in which the stationary property is required. Since these results are universal, they may provide many useful implications. In this paper, we will apply the above

asymptotic independence results to three high-dimensional hypothesis testing problems: one-sample mean test, two-sample mean test and the regression coefficient test.

1.2 High-Dimensional Hypothesis Testing

High-dimensional hypothesis testing is an important research area in modern statistics. It has been frequently used in many application fields, such as genomics, medical imaging, risk management and web search. The motivation of studying high-dimensional test is that traditional tests, such as the Hotelling T-squared test, do not work in general when the data dimension is larger than the sample size due to the singularity of sample covariance matrix. A nature way to amend this problem is replacing the sample covariance matrix appearing in the Hotelling T-squared test statistic with a nonsingular matrix, such as the identity matrix and the diagonal matrix of sample covariance matrix. In this way, for example, Srivastava (2009), Park and Ayyala (2013), Wang et al. (2015), Feng et al. (2016), Feng et al. (2015) and Feng et al. (2017) developed tests for one-sample mean problem, while Bai and Saranadasa (1996), Srivastava and Du (2008), Chen and Qin (2010), and Gregory et al. (2015) developed tests for two-sample mean problem. In addition, Goeman et al. (2006) and Lan et al. (2014), for instance, considered testing regression coefficients in high-dimensional linear models. All these tests are sum-type tests, based on the summation of parameter estimators. It is well known that the sum-type tests generally perform well as data are dense, i.e. most of the parameters are nonzero under the local alternative. However, it may be inefficient when data are sparse, where only a few parameters are nonzero under local alternative. To establish high-dimensional tests for sparse data, Cai et al. (2014), Zhong et al. (2013) and Chen et al. (2019) proposed some max-type tests, which typically perform well on sparse data, but worse when the data become dense.

In practice, it is often difficult to determine whether data are sparse or not. Thus, many efforts have been devoted to develop tests with good and robust performance under both data conditions. For example, Fan et al. (2015) proposed a power enhancement procedure by a screening technique for high-dimensional tests. They combined the power enhancement component with an asymptotically pivotal statistic to strengthen powers under sparse alternatives. Xu et al. (2016) initiated an adaptive test for high-dimensional two-sample mean test. It combines information across a class of sum-of-powers tests, including tests based on the sum-of-squares of the mean differences and the supremum mean difference. Wu et al. (2019) extended the adaptive test to generalized linear models. In He et al. (2021), the authors constructed U-statistics of different orders that are asymptotically independent of the max-type test statistics in high-dimensional tests, upon which an adaptive testing procedure is proposed. However, these results are based on Hsing (1995), which require data to be sampled from stationary and α-mixing random variables. In fact, the α-mixing property is hardly checked in practice, which greatly limits the application of these methods. In this paper, by using the novel asymptotic independence analysis between the sum and maximum aforementioned, we solve the problem without the stationary assumption or the α-mixing property and propose a series of high-dimensional tests including one-sample mean test, two-sample mean test and the regression coefficient test. Numerical results demonstrate strong robustness of the proposed tests regardless data being sparse or not.

The main contributions of this paper are listed as follows. (1) We establish the asymptotic distribution of the maximum of dependent Gaussian random variables under a general assumption. (2) We prove the asymptotic independence between the sum and maximum of dependent Gaussian random variables without the stationary or the α-mixing property. (3) We propose three high-dimensional combo-type tests based on the above asymptotic properties. They are one-sample mean test, two-sample mean test and the regression coefficient test. Numerical examples on simulated and real-world data demonstrate strong robustness of our tests, on both sparse and dense datasets.

The rest of the paper is organized as follows. In Section 2, we state our theoretical results, including the asymptotic distributions of the sum and maximum statistics, and the asymptotic independence between them. In Section 3, we propose a series of tests for high-dimensional data based on these theoretical results. Then, we demonstrate the simulation results of the proposed tests in comparison with some existing ones in Section 4, followed by application in Section 5. Finally, we present some concluding remarks in Section 6, while providing some extended results and technical proofs in the supplementary material.

2. Asymptotic Independence of Sum and Maximum of Dependent Random Variables

First, in this section, we study the asymptotic normality of the sum of dependent random variables. For each $p \geqslant 2$, let $Z_{p1}, \cdots, Z_{pp}$ be $N(0,1)$-distributed random variables with $p \times p$ covariance matrix $\mathbf{\Sigma}_p$. If there is no danger of confusion, we simply write "$Z_1, \cdots, Z_p$" for "$Z_{p1}, \cdots, Z_{pp}$" and

"$\mathbf{\Sigma}$" for "$\mathbf{\Sigma}_p$". The following assumption is needed:

$$\lim_{p\to\infty} \frac{[\mathrm{tr}(\mathbf{\Sigma}^{2+\delta})]^2}{[\mathrm{tr}(\mathbf{\Sigma}^2)]^{2+\delta}} = 0 \ \text{ for some } \delta > 0. \tag{2.1}$$

Assumption (2.1) with $\delta = 2$ is the same as condition (3.7) in Chen and Qin (2010), and here we make it more general. Although in applications the true covariance matrix $\mathbf{\Sigma}$ is usually unknown, this condition assures the practitioners that our results would be applicable to a wide range of problems. For instance, if all eigenvalues of $\mathbf{\Sigma}$ are bounded above and are bounded below from zero, it is trivial to see that (2.1) holds.

THEOREM 1. *Under Assumption* (2.1), $\frac{Z_1^2+\cdots+Z_p^2-p}{\sqrt{2tr(\mathbf{\Sigma}^2)}} \to N(0,1)$ *in distribution as* $p \to \infty$.

Theorem 1 shows that the sum of squares of the dependent Gaussian random variables has the asymptotic normality if the covariance matrix satisfies Assumption (2.1).

Next, for the same Gaussian random variables, we consider the asymptotic distribution of $\max_{1\leqslant i\leqslant p} Z_i^2$. Let $|A|$ denote the cardinality of the set A. The following assumption will be imposed:

Let $\mathbf{\Sigma} = (\sigma_{ij})_{1\leqslant i,j\leqslant p}$. For some $\varrho \in (0,1)$, assume $|\sigma_{ij}| \leqslant \varrho$ for all $1 \leqslant i < j \leqslant p$ and $p \geqslant 2$. Suppose $\{\delta_p;\, p\geqslant 1\}$ and $\{\kappa_p;\, p \geqslant 1\}$ are positive constants with $\delta_p = o(1/\log p)$ and $\kappa = \kappa_p \to 0$ as $p \to \infty$. For $1 \leqslant i \leqslant p$, define $B_{p,i} = \big\{1 \leqslant j \leqslant p;\, |\sigma_{ij}| \geqslant \delta_p\big\}$

and $C_p = \big\{1 \leqslant i \leqslant p;\ |B_{p,i}| \geqslant p^{\kappa}\big\}$. We assume that $|C_p|/p \to 0$ as $p \to \infty$. (2.2)

Theorem 2. *Suppose Assumption (2.2) holds. Then $\max_{1\leqslant i\leqslant p} Z_i^2 - 2\log p + \log\log p$ converges to a Gumbel distribution with cdf $F(x) = \exp\{-\frac{1}{\sqrt{\pi}}e^{-x/2}\}$ as $p \to \infty$.*

Remark 1. *Cai et al. (2014) obtained the above limiting distribution of $\max_{1\leqslant i\leqslant p} Z_i^2$ under the assumption that $\max_{1\leqslant i\leqslant p}\sum_{j=1}^{p}\sigma_{ij}^2 \leqslant C_0$ for each $p \geqslant 1$, where C_0 is a constant free of p. In the following we will see that their result is a special case of Theorem 2. In fact, let $\delta_p = (\log p)^{-2}$ for $p \geqslant e^e$, then for each $1 \leqslant i \leqslant p$, $\delta_p^2 \cdot |B_{p,i}| \leqslant \sum_{j=1}^{p}\sigma_{ij}^2 \leqslant C_0$. Hence, $|B_{p,i}| \leqslant C_0 \cdot (\log p)^2 < p^{\kappa}$ where $\kappa = \kappa_p := 5(\log\log p)/\log p$ for large p. As a result, $|C_p| = 0$, which implies the results of Theorem 2.*

A closely related but not exactly the same result by Fan and Jiang (2019) shows that $\delta_p = o(1/\log p)$ in Assumption (2.2) can not be relaxed. Their statistic is $\max_{1\leqslant i\leqslant p} Z_i$ in contrast to $\max_{1\leqslant i\leqslant p}|Z_i|$ here. We expect that $\delta_p = o(1/\log p)$ is also the critical threshold for $\max_{1\leqslant i\leqslant p}|Z_i|$.

Theorem 2 is proved by using the spirit of the proof of Lemma 6 from Cai et al. (2014). There are two purposes to derive the result. First, the conditions imposed in our theorem is weaker than those required in Lemma 6 from Cai et al. (2014), which has been discussed in Remark 1. This allows us to apply this type of results to a more general covariance matrix $\mathbf{\Sigma}$. Secondly, part of the steps in the proof of Theorem 2 will also be used in the proof of Theorem 3 stated next.

To proceed, we need more notations and an additional assumption. For two sequences of numbers $\{a_p \geqslant 0;\, p \geqslant 1\}$ and $\{b_p > 0;\, p \geqslant 1\}$, we write $a_p \ll b_p$ if $\lim_{p\to\infty} \frac{a_p}{b_p} = 0$. The following assumption will be used:

$$\text{There exist } C > 0 \text{ and } \varrho \in (0,1) \text{ so that } \max_{1\leqslant i<j\leqslant p} |\sigma_{ij}| \leqslant \varrho \text{ and } \max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} \sigma_{ij}^2 \leqslant (\log p)^C$$
$$\text{for all } p \geqslant 3;\ p^{-1/2}(\log p)^C \ll \lambda_{min}(\mathbf{\Sigma}) \leqslant \lambda_{max}(\mathbf{\Sigma}) \ll \sqrt{p}(\log p)^{-1} \text{ and}$$
$$\lambda_{max}(\mathbf{\Sigma})/\lambda_{min}(\mathbf{\Sigma}) = O(p^{\tau}) \text{ for some } \tau \in (0, 1/4). \tag{2.3}$$

Assumption (2.3) is actually stronger than both (2.1) and (2.2). To see this, assume (2.3) holds now. To derive (2.1), observe that $\mathrm{tr}(\mathbf{\Sigma}^{2+\delta}) \leqslant p \cdot \lambda_{max}(\mathbf{\Sigma})^{2+\delta}$ and $\mathrm{tr}(\mathbf{\Sigma}^2) \geqslant p \cdot \lambda_{min}(\mathbf{\Sigma})^2$. Then $\frac{[\mathrm{tr}(\mathbf{\Sigma}^{2+\delta})]^2}{[\mathrm{tr}(\mathbf{\Sigma}^2)]^{2+\delta}} \leqslant \frac{1}{p^{\delta}} \cdot \left(\frac{\lambda_{max}(\mathbf{\Sigma})}{\lambda_{min}(\mathbf{\Sigma})}\right)^{4+2\delta} = O\left(\frac{1}{p^{\delta-(4+2\delta)\tau}}\right) \to 0$ by choosing $\delta = 2$ and using the assumption $\tau \in (0, 1/4)$ stated in (2.3). We then get (2.1) with $\delta = 2$. To deduce (2.2), we replace "C_0" in Remark 1 with "$(\log p)^C$". By the same argument as that in Remark 1 and choosing $\delta_p = (\log p)^{-2}$, we see $|B_{p,i}| \leqslant C_0 \cdot (\log p)^{C+2} < p^{\kappa}$, where $\kappa = \kappa_p := (C+3)(\log\log p)/\log p$ for $p \geqslant e^e$. Hence, $|C_p| = 0$ and Assumption (2.2) holds.

Theorem 3. *Under Assumption* (2.3), *the following holds:* $\frac{Z_1^2+\cdots+Z_p^2-p}{\sqrt{2tr(\mathbf{\Sigma}^2)}}$

and $\max_{1\leqslant i\leqslant p} Z_i^2 - 2\log p + \log\log p$ *are asymptotically independent as* $p \to \infty$.

Importantly, notice that the above asymptotic independence result holds without the stationary assumption or the α-mixing condition. Regarding the assumption on the spectrum, in high-dimensional statistics literature, it is common to assume $[\lambda_{min}(\mathbf{\Sigma}), \lambda_{max}(\mathbf{\Sigma})] \subset [a, b]$, with $0 < a < b < \infty$. Note that this is stronger than our assumption on the eigenvalues of $\mathbf{\Sigma}$ in (2.3). In fact, Assumption (2.3) allows that the largest eigenvalue goes to infinity and the smallest eigenvalue goes to zero. Thus, Theorem 3 provides more general result and more freedom and practicality in application.

3. Application: High-Dimensional Testing Problems

In this section, we will apply the theoretical results derived in Section 2 to three high-dimensional testing problems: one-sample mean test, two-sample mean test and the regression coefficient test. The first test will be presented in the following subsections, while two-sample mean test and regression coefficient test will be presented in the supplementary material.

3.1 One-Sample Mean Test

Assume $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are independent and identically distributed p-dimensional random vectors from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The classical one-sample mean testing problem considers

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \mathbf{0}. \tag{3.4}$$

In the traditional setting where p is fixed, this topic is covered in classic textbooks on multivariate analysis such as in Anderson (2003), Eaton (1983) and Muirhead (1982). Starting from this century, a tremendous effort has been made for the test towards the high-dimensional setting, where both n and p go to infinity. In the following we will highlight part of these work en route to a problem we are interested in: the test (3.4) under the situation $n \leqslant p$. This is a typical problem of interest in high-dimensional statistics with small n and large p.

Let $\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$ and $\hat{\mathbf{S}} = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T$ be the sample mean and the sample covariance matrix of $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$, respectively. The Hotelling T^2-statistic is defined by $n\bar{\boldsymbol{X}}^T\hat{\mathbf{S}}^{-1}\bar{\boldsymbol{X}}$; see Hotelling (1931). For the case with $n > p$, Bai and Saranadasa (1996) studied the Hotelling statistic. When $n \leqslant p$, however, the matrix $\hat{\mathbf{S}}$ is no longer invertible, which motivates the design of new statistics. By replacing $\hat{\mathbf{S}}$ with its diagonal matrix in the Hotelling T^2-statistic, Srivastava and Du (2008) and Srivastava (2009) proposed a scale-invariant test for (3.4), defined by

$$T_{sum}^{(1)} = \frac{n\bar{\boldsymbol{X}}^T\hat{\mathbf{D}}^{-1}\bar{\boldsymbol{X}} - (n-1)p/(n-3)}{\sqrt{2[\mathrm{tr}(\hat{\mathbf{R}}^2) - p^2/(n-1)]}}, \tag{3.5}$$

where $\hat{\mathbf{D}}$ is the diagonal matrix of the sample covariance matrix $\hat{\mathbf{S}}$, and $\hat{\mathbf{R}} = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{S}}\hat{\mathbf{D}}^{-1/2}$ is the sample correlation matrix. The major ingredient of $T_{sum}^{(1)}$ can be written as a sum of random variables, so we sometimes call it a "sum-type" statistic. In general, the performance of sum-type statistics are not ideal in sparse cases when only a few entries in $\boldsymbol{\mu}$ in the sum are non-zero; see Cai et al. (2014) for more detailed discussion. Zhong et al. (2013) proposed two alternative tests by first thresholding two statistics based on the sample means and then maximizing over a range of thresholding levels. Denote $\bar{\boldsymbol{X}} = (\bar{\boldsymbol{X}}_1, \cdots, \bar{\boldsymbol{X}}_p)^T$. The L_2-version of the thresholding statistic is

$$T_{HC2} = \max_{s\in\mathcal{S}} \frac{T_{2n}(s) - \hat{\mu}(s)}{\hat{\sigma}(s)}, \tag{3.6}$$

where $\mathcal{S}$ is a subset of the interval $(0,1)$,

$$T_{2n}(s) = \sum_{j=1}^{p} n\left(\bar{\boldsymbol{X}}_j/\sigma_j\right)^2 I\left(|\bar{\boldsymbol{X}}_j| \geqslant \sigma_j\sqrt{\lambda_s/n}\right),$$

$$\hat{\mu}(s) = p\left\{2\lambda_p^{1/2}(s)\phi(\lambda_p^{1/2}(s)) + 2\bar{\Phi}(\lambda_p^{1/2}(s))\right\},$$

$$\hat{\sigma}^2(s) = p\left\{2\left[\lambda_p^{3/2}(s) + 3\lambda_p^{1/2}(s)\right]\phi(\lambda_p^{1/2}(s)) + 6\bar{\Phi}(\lambda_p^{1/2}(s))\right\}.$$

Here $\lambda_s(p) = 2s\log p$, and $\phi(\cdot)$, $\bar{\Phi}(\cdot)$ are the density and survival functions of the standard normal distribution, respectively. Fan et al. (2015) proposed a novel procedure by adding a power enhancement component which is asymptotically zero under the null and diverges under some specific regions of alternatives. Their test statistic is

$$J = J_0 + J_1, \tag{3.7}$$

where the power enhancement component J_0 is $J_0 = \sqrt{p}\sum_{j=1}^{p} \bar{\boldsymbol{X}}_j^2 \hat{\sigma}_j^{-2} I(|\bar{\boldsymbol{X}}_j| > \hat{\sigma}_j \delta_{p,n})$, and J_1 is the standard Wald statistic $J_1 = \frac{\bar{\boldsymbol{X}}^T \widehat{\mathrm{var}}^{-1}(\hat{\boldsymbol{X}})\bar{\boldsymbol{X}} - p}{2\sqrt{p}}$. Here $\hat{\sigma}_j^2$ is the sample variance of the jth coordinate of the population vector, $\delta_{p,n}$ is a thresholding parameter and $\widehat{\mathrm{var}}^{-1}(\hat{\boldsymbol{X}})$ is a consistent estimator of the asymptotic inverse covariance matrix of $\bar{\boldsymbol{X}}$. However, the power enhancement component would be negligible if the signal is not very strong. As we mentioned before, Cai et al. (2014) showed that extreme-value statistics are particularly powerful against sparse alternatives and possess certain optimal properties. Hence, we propose a statistic by compromising the sum-type statistic from (3.5) and an extreme-value statistic, based on our results in Section 2, which will be compared with aforementioned baselines numerically in Section 4.1. As will be confirmed later, our method performs very well regardless of the sparsity of the alternative hypothesis.

We now formally introduce our approach. Define

$$T_{max}^{(1)} = n \cdot \max_{1 \leqslant i \leqslant p} \frac{\bar{\boldsymbol{X}}_i^2}{\hat{\sigma}_{ii}^2}, \tag{3.8}$$

where $\bar{\boldsymbol{X}}_i$ is the ith coordinate of $\bar{\boldsymbol{X}} = \frac{1}{n}(\mathbf{X}_1 + \cdots \mathbf{X}_n) \in \mathbb{R}^p$ and $\hat{\sigma}_{ii}^2$ is the sample variance of the ith coordinate of the population vector, that is, if we write $\mathbf{X}_j = (x_{1j}, \cdots, x_{pj})^T$ for each $1 \leqslant j \leqslant n$, then $\hat{\sigma}_{ii}^2$ is the sample variance of the i.i.d. random variables $x_{i1}, x_{i2}, \cdots, x_{in}$. Firstly, the asymptotic distribution of $T_{max}^{(1)}$ will be presented which needs more notations. Let $\mathbf{R} = \mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2} = (\rho_{ij})_{1 \leqslant i,j \leqslant p}$ denote the population correlation matrix, where $\mathbf{D}$ is the diagonal matrix of $\boldsymbol{\Sigma}$. The following assumption will be imposed:

$$\text{There exists } \epsilon \in \left(\frac{1}{2}, 1\right] \text{ and } K > 1 \text{ such that } K^{-1}p^{\epsilon} \leqslant n \leqslant Kp^{\epsilon} \text{ and}$$
$$\sup_{p \geqslant 2} \frac{1}{p}\mathrm{tr}(\mathbf{R}^i) < \infty \text{ for } i = 2, 3, 4. \tag{3.9}$$

Note that (3.9) is the same as assumptions (3.1) and (3.2) from Srivastava (2009). If the eigenvalues of the correlation matrix $\mathbf{R}$ are bounded, the second condition of (3.9) will hold automatically. For rigor of mathematics, we assume n depends on p and sometimes write n_p when there is a possible confusion.

Theorem 4. *Under the null hypothesis in* (3.4)*, the following holds as* $p \to \infty$*:*

(i) If (3.9) *holds, then* $T_{sum}^{(1)} \to N(0,1)$ *in distribution;*

(ii) If (2.2) *holds with "*$\mathbf{\Sigma}$*" replaced by "*$\mathbf{R}$*" and* $\log p = o(n^{1/3})$*, then* $T_{max}^{(1)} - 2\log p + \log\log p$ *converges weakly to a Gumbel distribution with cdf* $F(x) = \exp\{-\frac{1}{\sqrt{\pi}}\exp(-x/2)\}$*;*

(iii) Assume (3.9) *is true. If* (2.3) *holds with "*$\mathbf{\Sigma}$*" replaced by "*$\mathbf{R}$*", then* $T_{sum}^{(1)}$ *and* $T_{max}^{(1)} - 2\log p + \log\log p$ *are asymptotically independent.*

Part (i) of the above theorem is from Srivastava (2009), which is also a corollary of the recent work by Jiang and Li (2021). For the sum-type test, a level-α test will be performed through rejecting H_0 when $T_{sum}^{(1)}$ is larger than the $(1-\alpha)$-quantile $z_\alpha = \Phi^{-1}(1-\alpha)$ where $\Phi(y)$ is the cdf of $N(0,1)$. For the max-type test, a level-α test will then be performed through rejecting H_0 when $T_{max}^{(1)} - 2\log p + \log\log p$ is larger than the $(1-\alpha)$-quantile $q_\alpha = -\log\pi - 2\log\log(1-\alpha)^{-1}$ of the Gumbel distribution $F(x)$.

Based on Theorem 4, we propose a combo-type test statistic by combining the max-type and the sum-type tests. It is defined by

$$T_{com}^{(1)} = \min\{P_S^{(1)}, P_M^{(1)}\}, \tag{3.10}$$

where $P_S^{(1)} = 1 - \Phi\left\{T_{sum}^{(1)}\right\}$ and $P_M^{(1)} = 1 - F(T_{max}^{(1)} - 2\log p + \log\log p)$. Note that $P_S^{(1)}$ and $P_M^{(1)}$ are the p-values for the tests by using statistics $T_{sum}^{(1)}$ and $T_{max}^{(1)}$, separately, and $T_{com}^{(1)}$ is defined by the smaller one, whose asymptotic distribution can be characterized by the minimum of two standard uniform random variables.

Corollary 1. *Assume the conditions in Theorem 4(iii) hold. Then* $T_{com}^{(1)}$ *from* (3.10) *converges weakly to a distribution with density* $G(w) = 2(1-w)I(0 \leqslant w \leqslant 1)$ *as* $p \to \infty$.

According to Corollary 1, the proposed combo-type test allows us to perform a level-α test by rejecting the null hypothesis when $T_{com}^{(1)} < 1 - \sqrt{1-\alpha} \approx \frac{\alpha}{2}$ as α is small. We now discuss the power functions. First, the power function of our combo-type test is

$$\begin{aligned} \beta_C^{(1)}(\boldsymbol{\mu}, \alpha) &= P\left(T_{com}^{(1)} < 1 - \sqrt{1-\alpha}\right) = P\left(P_M^{(1)} < 1 - \sqrt{1-\alpha} \text{ or } P_S^{(1)} < 1 - \sqrt{1-\alpha}\right) \\ &\geqslant \max\left\{P\left(P_S^{(1)} < 1 - \sqrt{1-\alpha}\right), P\left(P_M^{(1)} < 1 - \sqrt{1-\alpha}\right)\right\} \\ &\approx \max\left\{\beta_S^{(1)}(\boldsymbol{\mu}, \alpha/2), \beta_M^{(1)}(\boldsymbol{\mu}, \alpha/2)\right\} \end{aligned} \tag{3.11}$$

when α is small, where $\beta_M^{(1)}(\boldsymbol{\mu}, \alpha)$ and $\beta_S^{(1)}(\boldsymbol{\mu}, \alpha)$ are the power functions of $T_{max}^{(1)}$ and $T_{sum}^{(1)}$ with significant level α, respectively. From Srivastava (2009), the power function of $T_{sum}^{(1)}$ is

$$\beta_S^{(1)}(\boldsymbol{\mu}, \alpha) = \lim_{p\to\infty} \Phi\left(-z_\alpha + \frac{n\boldsymbol{\mu}^T\mathbf{D}^{-1}\boldsymbol{\mu}}{\sqrt{2\text{tr}(\mathbf{R}^2)}}\right), \tag{3.12}$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$ is the $(1-\alpha)$-quantile of $N(0,1)$. Due to (3.11), we have $\beta_C^{(1)}(\boldsymbol{\mu}, \alpha) \geqslant \lim_{p\to\infty} \Phi\left(-z_{\alpha/2} + \frac{n\boldsymbol{\mu}^T\mathbf{D}^{-1}\boldsymbol{\mu}}{\sqrt{2\text{tr}(\mathbf{R}^2)}}\right)$. Denote $\mathbf{D} = \text{diag}(\sigma_{11}^2, \cdots, \sigma_{pp}^2)$.

By the same argument at that from Theorem 2 in Cai et al. (2014), the asymptotic power of $T_{max}^{(1)}$ converges to one if $\max_{1\leqslant i\leqslant p}|\mu_i/\sigma_{ii}| \geqslant c\sqrt{\log p/n}$ for a certain constant c, and also the nonzero μ_i are randomly uniformly sampled with sparsity level $\gamma < 1/4$, i.e., the number of nonzero μ_i is less than $p^\gamma, \gamma < 1/4$. Thus, according to (3.11), the power function of our proposed test $T_{com}^{(1)}$ also converges to one in this case. Similarly, according to Theorem 3 in Cai et al. (2014), the condition $\max_{1\leqslant i\leqslant p}|\mu_i/\sigma_{ii}| \geqslant c\sqrt{\log p/n}$ is minimax rate optimal for testing against sparse alternatives. If c is sufficiently small, then any α-level test is unable to reject the null hypothesis with probability tending to one. It is shown in Cai et al. (2014) that $T_{max}^{(1)}$ enjoys a certain optimality against sparse alternatives. By (3.11), our test $T_{com}^{(1)}$ also has this optimality.

In order to get a rough picture of the asymptotic power comparison between $T_{sum}^{(1)}, T_{max}^{(1)}$ and $T_{com}^{(1)}$, now we simply assume that $\mathbf{\Sigma} = \mathbf{I}_p$. There are m nonzeros μ_i and they are all equal to $\delta \neq 0$. Equation (3.12) gives $\beta_S^{(1)}(\boldsymbol{\mu}, \alpha) = \lim_{p\to\infty} \Phi\left(-z_\alpha + \frac{nm\delta^2}{\sqrt{2p}}\right)$.

We consider two special cases:

(1) *Dense case*: $\delta = O(n^{-\xi})$ and $m = O(p^{1/2}n^{2\xi-1})$ with $\xi \in (1/2, 5/6]$. We also assume $\log p = o(n^{\xi-\frac{1}{2}})$, hence $\log p = o(n^{1/3})$. As a consequence, the requirement on p vs n imposed in Theorem 4(ii) is fulfilled. Obvi-

ously, the number of nonzero μ_i goes to infinity. The power function for $T_{max}^{(1)}$ is given by $\beta_M^{(1)}(\boldsymbol{\mu}, \alpha) = P\left(T_{max}^{(1)} - 2\log p + 2\log\log p > q_\alpha\right)$. In this case, we will show in Section S3.4 of the supplementary material that $\beta_M^{(1)}(\boldsymbol{\mu}, \alpha) \approx \alpha$, which means that $T_{max}^{(1)}$ is not effective or useful. Consequently, we have $\beta_C^{(1)}(\boldsymbol{\mu}, \alpha) \approx \beta_S^{(1)}(\boldsymbol{\mu}, \alpha/2)$. When the significant level α is small, the difference between $\beta_S^{(1)}(\boldsymbol{\mu}, \alpha)$ and $\beta_S^{(1)}(\boldsymbol{\mu}, \alpha/2)$ is negligible. So our proposed test $T_{com}^{(1)}$ has similar performance as $T_{sum}^{(1)}$ in this dense case.

(2) *Sparse case*: $\delta = c\sqrt{\log p/n}$ for sufficient large constant c and $m = o((\log p)^{-1}p^{1/2})$. Here the value of m is much smaller than that in (1) and hence confirms the notion of "sparse". In this case, $\frac{nm\delta^2}{\sqrt{2p}} \to 0$, so $\beta_S^{(1)}(\boldsymbol{\mu}, \alpha) \approx \alpha$ and $T_{sum}^{(1)}$ is not effective or useful. Yet, $\beta_M^{(1)}(\boldsymbol{\mu}, \alpha) \to 1$ by an argument similar to Theorem 2 from Cai et al. (2014) as discussed above, which also leads to $\beta_C^{(1)}(\boldsymbol{\mu}, \alpha) \to 1$ in this sparse case. Additionally, in Fan et al. (2015) the quantity $\delta_{p,n}$ is chosen to be $\log\log n\sqrt{\log p/n}$, which implies that the screening set $\{i : \sqrt{n}|\bar{\boldsymbol{X}}_i| > \log\log n\sqrt{\log p}\}$ would be empty as probability tending to one. Thus, the power enhancement component of Fan et al. (2015) would be negligible in this case, which makes the standardized Wald test statistic the same as $T_{sum}^{(1)}$ since $\boldsymbol{\Sigma} = \mathbf{I}_p$. That is, their test is also

ineffective in this sparse case.

The above theoretical results and analysis, together with the simulation in the next section, indicate that our proposed test $T_{com}^{(1)}$ performs very well regardless of the sparsity of the alternative hypothesis, which is more convenient to use in various practical scenarios.

Due to space limitation, we will present the combo-type two-sample mean test and regression coefficient test as well as their simulation results in the supplementary material.

4. Simulation Results

In this section, we carry out a series of simulation study on the testing problems studied in the previous section, to compare different test statistics and validate the advantage of the proposed combo-type tests.

4.1 One-Sample Test Problem

Firstly, we conduct numerical examples on the one-sample test problem. We compare our combo-type test T_{com} in (3.10) (abbreviated as COM) with the sum-type test $T_{sum}^{(1)}$ in (3.5) by Srivastava (2009) (abbreviated as SUM), the max-type test $T_{max}^{(1)}$ in (3.8) (abbreviated as MAX), the Higher Criticism test T_{HC2} from (3.6) by Zhong et al. (2013) (abbreviated as HC2) and the power enhancement test J from (3.7) by Fan et al. (2015) (abbreviated as FLY). The dataset is simulated as follows.

EXAMPLE 1. We consider $\boldsymbol{X}_i = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{z}_i$ for $i = 1\cdots,n$, and each component of $\boldsymbol{z}_i$ is independently generated from three distributions: (1) the normal distribution $N(0,1)$; (2) the t distribution $t(3)/\sqrt{3}$; (3) the mixture normal random variable $V/\sqrt{1.8}$, where V has density function $0.1f_1(x) + 0.9f_2(x)$ with $f_1(x)$ and $f_2(x)$ being the densities of $N(0,9)$ and $N(0,1)$, respectively. We will work on two different sample sizes with $n = 100, 200$ and three different dimensions with $p = 200, 400, 600$. Under the null hypothesis, we set $\boldsymbol{\mu} = \mathbf{0}$ and the significance level $\alpha = 0.05$. The following three scenarios of covariance matrices will be considered.

(I) AR(1) model: $\boldsymbol{\Sigma} = (0.5^{|i-j|})_{1\leqslant i,j\leqslant p}$.

(II) $\boldsymbol{\Sigma} = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$ with $\mathbf{D} = \mathrm{diag}(\sigma_1^2,\cdots,\sigma_p^2)$ and $\mathbf{R} = \mathbf{I}_p + \boldsymbol{b}\boldsymbol{b}^T - \check{\mathbf{B}}$, where σ_i^2 are generated independently from $Uniform(1,2)$, $\boldsymbol{b} = (b_1,\cdots,b_p)^T$ and $\check{\mathbf{B}} = \mathrm{diag}(b_1^2,\cdots,b_p^2)$. The first $[p^{0.3}]$ entries of $\boldsymbol{b}$ are independently sampled from $Uniform(0.7,0.9)$, and the remaining entries are set to be zero, where $[\cdot]$ denotes taking integer part.

(III) $\boldsymbol{\Sigma} = \boldsymbol{\gamma}\boldsymbol{\gamma}^T + (\mathbf{I}_p - \rho_\epsilon \boldsymbol{W})^{-1}(\mathbf{I}_p - \rho_\epsilon \boldsymbol{W}^T)^{-1}$, where $\boldsymbol{\gamma} = (\gamma_1,\cdots,\gamma_{[p^{\delta_\gamma}]},0,0,\cdots,0)^T$. Here γ_i with $i = 1,\cdots,[p^{\delta_\gamma}]$ are generated independently from $Uniform(0.7,0.9)$.

Table 1: Sizes of tests for Example 1 with Scenario (I), $\alpha = 0.05$.

	Distribution	(1)			(2)			(3)		
	p	200	400	600	200	400	600	200	400	600
$n = 100$	MAX	0.053	0.062	0.082	0.026	0.052	0.045	0.044	0.039	0.061
	SUM	0.064	0.064	0.060	0.052	0.050	0.059	0.063	0.058	0.064
	COM	0.063	0.069	0.059	0.040	0.059	0.055	0.056	0.047	0.061
	HC2	0.028	0.044	0.034	0.033	0.029	0.032	0.038	0.025	0.044
	FLY	0.014	0.009	0.004	0.003	0.003	0.002	0.025	0.018	0.014
$n = 200$	MAX	0.046	0.060	0.049	0.045	0.041	0.045	0.042	0.045	0.032
	SUM	0.065	0.068	0.058	0.053	0.057	0.062	0.056	0.054	0.056
	COM	0.056	0.068	0.048	0.042	0.047	0.052	0.043	0.050	0.039
	HC2	0.019	0.027	0.030	0.031	0.024	0.023	0.029	0.020	0.029
	FLY	0.005	0.000	0.000	0.003	0.000	0.000	0.017	0.012	0.005

Let $\rho_\epsilon = 0.5$ and $\delta_\gamma = 0.3$. Let $\boldsymbol{W} = (w_{i_1 i_2})_{1 \leqslant i_1, i_2 \leqslant p}$ have a so-called rook form, i.e., all elements of $\boldsymbol{W}$ are zero except that $w_{i_1+1,i_1} = w_{i_2-1,i_2} = 0.5$ for $i_1 = 1, \cdots, p-2$ and $i_2 = 3, \cdots, p$, and $w_{1,2} = w_{p,p-1} = 1$.

Tables 1, 2, 3 report the empirical sizes of the five tests. SUM, MAX and COM can control the empirical sizes very well in most cases. However, the empirical sizes of HC2 and FLY can be much smaller than the nominal level in some cases.

Table 2: Sizes of tests for Example 1 with Scenario (II), $\alpha = 0.05$.

	Distribution	(1)			(2)			(3)		
	p	200	400	600	200	400	600	200	400	600
$n = 100$	MAX	0.058	0.070	0.065	0.044	0.037	0.039	0.048	0.042	0.047
	SUM	0.053	0.067	0.056	0.054	0.052	0.048	0.054	0.055	0.045
	COM	0.055	0.057	0.061	0.054	0.044	0.040	0.043	0.047	0.047
	HC2	0.022	0.011	0.013	0.005	0.015	0.005	0.011	0.011	0.006
	FLY	0.022	0.011	0.011	0.013	0.010	0.006	0.024	0.015	0.007
$n = 200$	MAX	0.053	0.054	0.076	0.025	0.042	0.025	0.044	0.040	0.041
	SUM	0.053	0.057	0.060	0.053	0.051	0.052	0.055	0.065	0.060
	COM	0.058	0.061	0.066	0.037	0.045	0.044	0.043	0.053	0.055
	HC2	0.003	0.011	0.006	0.010	0.006	0.003	0.004	0.005	0.008
	FLY	0.037	0.033	0.025	0.030	0.022	0.011	0.032	0.026	0.015

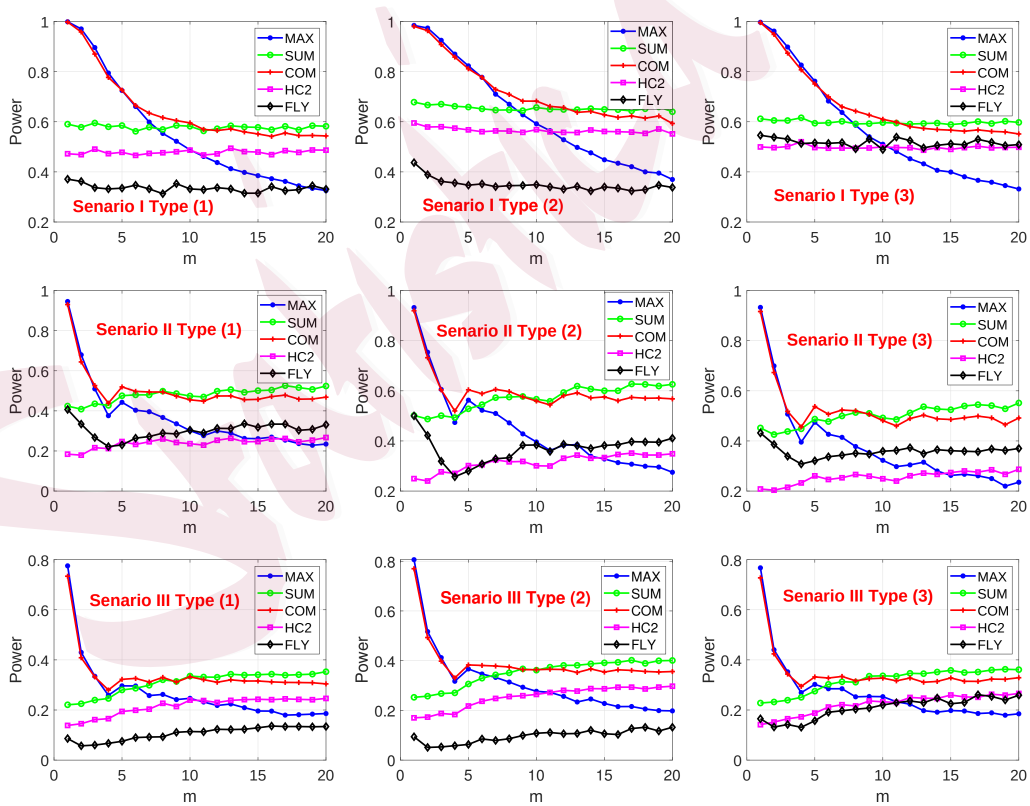
Next, we examine the power of each test. Our simulation shows that the power comparisons are similar for any combination of (n, p) with $n = 100, 200$ and $p = 200, 400, 600$. Hence, we present the case $n = 100$ and $p = 200$ for conciseness. Define $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)^T$. For different number of nonzero-mean variables $m = 1, \cdots, 20$, we consider $\mu_j = \delta$ for $0 < j \leqslant m$ and $\mu_j = 0$ for $j > m$. The parameter δ is chosen as $||\boldsymbol{\mu}||^2 = m\delta^2 = 0.5$. Figure 1 reports the power of the five tests. The power of MAX decreases as the number of nonzero-mean variables increases, which is as expected

Table 3: Sizes of tests for Example 1 with Scenario (III), $\alpha = 0.05$.

	Distribution	(1)			(2)			(3)		
	p	200	400	600	200	400	600	200	400	600
$n = 100$	MAX	0.054	0.066	0.059	0.053	0.040	0.033	0.049	0.039	0.043
	SUM	0.052	0.055	0.059	0.053	0.048	0.060	0.059	0.064	0.061
	COM	0.053	0.066	0.059	0.053	0.050	0.040	0.062	0.046	0.051
	HC2	0.034	0.038	0.035	0.032	0.030	0.025	0.036	0.030	0.030
	FLY	0.013	0.003	0.005	0.013	0.001	0.000	0.020	0.013	0.010
$n = 200$	MAX	0.053	0.058	0.063	0.034	0.027	0.038	0.049	0.039	0.050
	SUM	0.061	0.065	0.062	0.044	0.058	0.068	0.063	0.058	0.057
	COM	0.065	0.075	0.069	0.033	0.048	0.047	0.059	0.051	0.053
	HC2	0.035	0.032	0.032	0.019	0.029	0.019	0.029	0.023	0.024
	FLY	0.001	0.001	0.000	0.004	0.001	0.000	0.016	0.011	0.002

because, generally speaking, the max-type test is more powerful in sparse case and less powerful in non-sparse case. The power of SUM slightly increases with m and is higher than the power of HC2 and FLY in all cases. The proposed COM is as powerful as MAX when the number of variables with nonzero means is small (sparse case), and almost has the same power as SUM when the number of variables with nonzero means grows. In general, COM possesses the advantages of both MAX (in sparse case) and SUM (in non-sparse case), and outperforms HC2 and FLY in all scenarios. Observe that all the tests, except for COM, favor either the sparse or non-sparse case. Since in practice it is hard to justify whether the true underlying model is sparse or not, our proposed COM test, with its strong robustness, should be a more favorable choice over the competing approaches.

Figure 1: Power vs. number of variables with nonzero means for Example 4.1. The x-lab m denotes the number of variables with non-zero means; the y-lab is the empirical power.



5. Real Data Application

In this section, we further apply the results and test statistics obtained in Section 3 to two real data: a US stock data (dense model) and a search engine data (sparse model) in the supplementary material. As will be seen, the proposed combo-type test, COM, performs well on both datasets. Thus, it could serve as a "universal" test in practice no matter the true model is sparse or not.

5.1 US Stock Data

We apply the methods developed for one-sample mean test in Section 3.1 to a pricing problem in finance. Specifically, we investigate how financial returns of assets are related to their risk-free returns. Let $X_{ij} = R_{ij} - \mathrm{rf}_i$ denote the excess return of the jth asset at time i for $i = 1, \cdots, n$ and $j = 1, \cdots, p$, where R_{ij} is the return on asset j during period i and rf_i is the risk-free return rate of all assets during period i. We study the following pricing model

$$X_{ij} = \mu_j + \xi_{ij}, \tag{5.13}$$

for $i = 1, \cdots, n$ and $j = 1, \cdots, p$, or, in vector form, $\boldsymbol{X}_i = \boldsymbol{\mu} + \boldsymbol{\xi}_i$, where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^T$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$, and $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{ip})^T$ is the zero-mean error vector. The pricing model in (5.13) is the zero-factor model within the well known Arbitrage Pricing Theory (Ross, 1976), where "zero-factor" means that no additional factor is used to model the price. A common null hypothesis to be considered under the pricing model (5.13) is $H_0 : \boldsymbol{\mu} = \mathbf{0}$, which means that the excess return of any asset is zero on average, i.e. the return rate of any asset R_{ij} is equal to the risk-free return rate rf_i on average.

We consider the monthly return rates of the stocks that constitute the S&P 500 index over the period from January 2005 to November 2018. Since the stocks that made up the index changed over time and some stocks were created during this period, we only consider a total of 374 stocks that were included in the index during the entire time range. Figure 2 shows the sample mean of each stock in this period. We observe that most average returns are positive. In fact, as we enlarge the time range (increasing sample size n), the p-values of MAX, SUM and COM are eventually smaller than 0.05. These results suggest rejection of the null that the asset return does not only comes from the risk-free rates (on average), which is consistent with the views of many economists (Fama and French, 1993, 2015).

We further evaluate the tests by a random sampling procedure. Specifically, we randomly choose n samples from the whole dataset and apply

Figure 2: Histogram of sample means of stock monthly return rates in S&P 500.

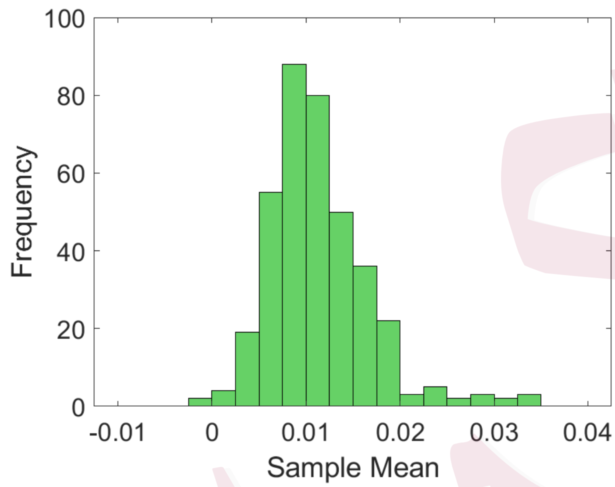


Table 4: Rejecting rates of each test in US stock Data.

	MAX	SUM	COM
$n = 30$	0.35	0.39	0.40
$n = 50$	0.40	0.51	0.51
$n = 70$	0.44	0.67	0.62
$n = 100$	0.52	0.86	0.83

MAX, SUM and COM on this new sample. For each n, we repeat this experiment for 1000 times. Table 4 reports the rejecting rates for each method with different n. From Table 4, we observe that SUM outperforms MAX in all cases by providing higher rejection rates. This is not surprising because for this data, the number of variables with nonzero means (assets with non-zero expected excess return) might be large, which is the case where sum-type tests could typically perform better than max-type tests. On the other hand, the combo-type test COM performs similarly as SUM overall. Therefore, COM does not lose efficiency in this problem.

6. Concluding Remarks

In this paper, we prove the asymptotic independence between the sum and maximum of dependent random variables without stationary assumptions or strong mixing conditions. Then we apply our results to high-dimensional testing problems. Our proposed combo-type tests perform well regardless data being sparse or not. Now we make some comments.

1. The normal assumption is essential in the proof of asymptotic independence. Hence, we assume the Gaussian assumption in the high-dimensional test problems. In some recent studies, such as Liang et al. (2019) and Chen and Xia (2021), some high-dimensional normality tests have been developed to check whether a p-dimensional random vector with large p is a Gaussian vector. In literature, we may not need the Gaussian assumption to analyze the asymptotic distribution of the sum-type and max-type test statistics, e.g., Cai et al. (2014); Chen and Qin (2010).

To prove the asymptotic independence between the sum and maximum of non-normal dependent random variables deserves further investigation.

2. To obtain the asymptotic distribution of the sum and maximum of dependent random variables, we assume the correlations between the random variables are not very strong. Recently, there has been much literature that consider high-dimensional testing problems without the weak correlation assumption, such as Wang and Xu (2021); Zhang et al. (2020). The analogue of our asymptotic independence result between the sum and maximum of dependent random variables with arbitrary covariance structures is also a very interesting and challenging problem.

3. The asymptotic independence results in Theorem 3 is universal. We believe it can be generalized and applied to many other applications, such as change point detection and statistical process controls.

Supplementary Materials

In the supplementary material, we propose the combo-type two-sample mean test and regression coefficient test and present corresponding simulation results to demonstrate the advantages of our proposed tests. The supplemental file also includes the technical proofs of our theoretical results.

Acknowledgement

We thank Professor Wei Xiong for very helpful discussions. The authors in this paper are listed in alphabetical order. The work of Xiaoyun Li was conducted when the author worked at Baidu research. The work of Tiefeng Jiang was conducted when the author was a visiting professor at Baidu research. Long Feng was partially supported by National Natural Science Foundation of China under Grant No. 12271271, the Fundamental Research Funds for the Central Universities under Grant No. ZB22000105. Binghui Liu was partially supported by the National Natural Science Foundation of China grant 12171079 and the National Key R&D Program of China grant 2020YFA0714102.

References

Anderson, C. and K. Turkman (1991). The joint limiting distribution of sums and maxima of stationary sequences. *Journal of Applied Probability 28*(1), 33–44.

Anderson, C. and K. Turkman (1993). Limiting joint distributions of sums and maxima in a statistical context. *Theory of Probability & Its Applications 37*(2), 314–316.

Anderson, C. and K. Turkman (1995). Sums and maxima of stationary sequences with heavy tailed distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, 1–10.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* John Wiley &

Sons.

Bai, Z. and H. Saranadasa (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 311–329.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys 2*, 107–144.

Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 349–372.

Chen, H. and Y. Xia (2021). A normality test for high-dimensional data based on the nearest neighbor approach. *Journal of the American Statistical Association*, Online published.

Chen, S. X., J. Li, and P.-S. Zhong (2019). Two-sample and anova tests for high dimensional means. *The Annals of Statistics 47*(3), 1443–1474.

Chen, S. X. and Y.-L. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics 38*(2), 808–835.

Chow, T. L. and J. L. Teugels (1978). The sum and the maximum of iid random variables. In *Proceedings of the 2nd Prague Symposium on Asymptotic Statistics*, Volume 45, pp. 394–403.

Davis, R. A. and T. Hsing (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability*, 879–917.

Eaton, M. L. (1983). Multivariate statistics: a vector space approach. *JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, USA, 1983, 512*.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of financial economics 116*(1), 1–22.

Fan, J. and T. Jiang (2019). Largest entries of sample correlation matrices from equi-correlated normal populations. *The Annals of Probability 47*(5), 3321–3374.

Fan, J., Y. Liao, and J. Yao (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica 83*(4), 1497–1541.

Feng, L., C. Zou, and Z. Wang (2016). Multivariate-sign-based high-dimensional tests for the two-sample location problem. *Journal of the American Statistical Association 111*(514), 721–735.

Feng, L., C. Zou, Z. Wang, and L. Zhu (2015). Two-sample behrens-fisher problem for high-dimensional data. *Statistica Sinica*, 1297–1312.

Feng, L., C. Zou, Z. Wang, and L. Zhu (2017). Composite t 2 test for high-dimensional data. *Statistica Sinica*, 1419–1436.

Goeman, J. J., S. A. Van De Geer, and H. C. Van Houwelingen (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical*

Methodology) 68(3), 477–493.

Gregory, K. B., R. J. Carroll, V. Baladandayuthapani, and S. N. Lahiri (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association 110*(510), 837–849.

He, Y., G. Xu, C. Wu, and W. Pan (2021). Asymptotically independent u-statistics in high-dimensional testing. *The Annals of Statistics 49*(1), 154–181.

Ho, H.-C. and T. Hsing (1996). On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *Journal of Applied Probability 33*(1), 138–145.

Ho, H.-C. and W. P. McCormick (1999). Asymptotic distribution of sum and maximum for gaussian processes. *Journal of Applied Probability 36*(4), 1031–1044.

Hotelling, H. (1931). The generalization of student's ratio. *The Annals of Mathematical Statistics 2*(3), 360–378.

Hsing, T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *The Annals of Probability*, 938–947.

Jiang, T. and P. Li (2021). Mean test with fewer observation than dimension and ratio unbiased estimator for correlation matrix. *arXiv preprint arXiv:2108.06892*.

Lan, W., H. Wang, and C.-L. Tsai (2014). Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics 66*(2), 279–301.

Liang, J., M.-L. Tang, and X. Zhao (2019). Testing high-dimensional normality based on classical skewness and kurtosis with a possible small sample size. *Communications in Statistics - Theory and Methods 48*(23), 5719–5732.

McCormick, W. and Y. Qi (2000). Asymptotic distribution for the sum and maximum of gaussian processes. *Journal of Applied Probability 37*(4), 958–971.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory.* John Wiley & Sons.

Park, J. and D. N. Ayyala (2013). A test for the mean vector in large dimension and small samples. *Journal of Statistical Planning and Inference 143*(5), 929–943.

Peng, Z. and S. Nadarajah (2003). On the joint limiting distribution of sums and maxima of stationary normal sequence. *Theory of Probability & Its Applications 47*(4), 706–709.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis 100*(3), 518–532.

Srivastava, M. S. and M. Du (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis 99*(3), 386–402.

Wang, L., B. Peng, and R. Li (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association 110*(512), 1658–1669.

Wang, R. and W. Xu (2021). An approximate randomization test for high-dimensional two-sample behrens-fisher problem under arbitrary covariances. *arXiv preprint* *arXiv:2108.01860*.

Wu, C., G. Xu, and W. Pan (2019). An adaptive test on high-dimensional parameters in generalized linear models. *Statistica Sinica 29*(4), 2163–2186.

Xu, G., L. Lin, P. Wei, and W. Pan (2016). An adaptive two-sample test for high-dimensional means. *Biometrika 103*(3), 609–624.

Zhang, J.-T., J. Guo, B. Zhou, and M.-Y. Cheng (2020). A simple two-sample test in high dimensions based on l 2-norm. *Journal of the American Statistical Association 115*(530), 1011–1027.

Zhong, P.-S., S. X. Chen, and M. Xu (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *The Annals of Statistics 41*(6), 2820–2851.

School of Statistics and Data Science, KLMDASR, LEBPS, and LPMC, Nankai University

E-mail: flnankai@nankai.edu.cn

School of Statistics, University of Minnesota

E-mail: jiang040@umn.edu

LinkedIn Corporation

E-mail: xl374@scarletmail.rutgers.edu

School of Mathematics and Statistics & KLAS, Northeast Normal University

E-mail: liubh100@nenu.edu.cn