The Annals of Statistics
2013, Vol. 41, No. 6, 2820–2851
DOI: 10.1214/13-AOS1168
© Institute of Mathematical Statistics, 2013

TESTS ALTERNATIVE TO HIGHER CRITICISM FOR HIGH-DIMENSIONAL MEANS UNDER SPARSITY AND COLUMN-WISE DEPENDENCE

BY PING-SHOU ZHONG[1], SONG XI CHEN[2,4] AND MINYA XU[3,4]

Michigan State University, Peking University and Iowa State University, and Peking University

We consider two alternative tests to the Higher Criticism test of Donoho and Jin [*Ann. Statist.* **32** (2004) 962–994] for high-dimensional means under the sparsity of the nonzero means for sub-Gaussian distributed data with unknown column-wise dependence. The two alternative test statistics are constructed by first thresholding L_1 and L_2 statistics based on the sample means, respectively, followed by maximizing over a range of thresholding levels to make the tests adaptive to the unknown signal strength and sparsity. The two alternative tests can attain the same detection boundary of the Higher Criticism test in [*Ann. Statist.* **32** (2004) 962–994] which was established for uncorrelated Gaussian data. It is demonstrated that the maximal L_2-thresholding test is at least as powerful as the maximal L_1-thresholding test, and both the maximal L_2 and L_1-thresholding tests are at least as powerful as the Higher Criticism test.

1. Introduction.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent and identically distributed (I.I.D.) p-variate random vectors generated from the following model:

$$\mathbf{X}_i = \mathbf{W}_i + \boldsymbol{\mu} \qquad \text{for } i = 1, \ldots, n, \tag{1.1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$ is a p-dimensional unknown vector of means, $\mathbf{W}_i = (W_{i1}, \ldots, W_{ip})^T$ and $\{\mathbf{W}_i\}_{i=1}^n$ are I.I.D. random vectors with zero mean and common covariance $\boldsymbol{\Sigma}$. For the ith sample, $\{W_{ij}\}_{j=1}^p$ is a sequence of weakly stationary dependent random variables with zero mean and variances σ_j^2. Motivated by the high-dimensional applications arising in genetics, finance and other fields, the cur-

Received March 2013; revised September 2013.

[1]Supported in part by NSF Grants DMS-13-09156 and DMS-12-09112.

[2]Supported in part by NSF Grant DMS-13-09210 and NSFC Key Grant 10901010 and Grant 71371016.

[3]Supported in part by NSFC Key Grant 10901010 and the Center for Statistical Science at Peking University.

[4]Supported in part by Key Laboratory of Mathematical Economics and Quantity Finance of MOE at Peking University.

MSC2010 subject classifications. Primary 62H15; secondary 62G20, 62G32.

Key words and phrases. Large deviation, large p, small n, optimal detection boundary, sparse signal, thresholding, weak dependence.

rent paper focuses on testing high-dimensional hypotheses

$$H_0: \boldsymbol{\mu} = 0 \quad \text{vs} \quad H_1: \text{nonzero } \mu_j \text{ are sparse and faint.} \tag{1.2}$$

The specifications for the sparsity and faintness in the above H_1 are the following. There are $p^{1-\beta}$ nonzero μ_j's (signals) for a $\beta \in (1/2, 1)$, which are sparse since the signal bearing dimensions constitute only a small fraction of the total p dimensions. Also under the H_1, the signal strength is faint in that the nonzero $\mu_j = \sqrt{2r \log(p)/n}$ for $r \in (0, 1)$. These specification of the H_1 have been the most challenging "laboratory" conditions in developing novel testing procedures under high dimensionality.

Donoho and Jin (2004) pioneered the theory of the Higher Criticism (HC) test which was originally conjectured in Tukey (1976), and showed that the HC test can attain the optimal detection boundary established by Ingster (1997) for uncorrelated Gaussian random vectors ($\boldsymbol{\Sigma} = \mathbf{I}_p$). The optimal detection boundary is a phase-diagram in the space of (β, r), the two quantities which define the sparsity and the strength of nonzero μ_j's under the H_1, such that if (β, r) lies above the boundary, there exists a test which has asymptotically diminishing probabilities of the type I and type II errors simultaneously; and if (β, r) is below the boundary, no such test exists. Hall and Jin (2008, 2010) investigated the impacts of the column-wise dependence on the HC test. In particular, Hall and Jin (2008) found that the HC test is adversely affected if the dependence is of long range dependent. If the dependence is weak, and the covariance matrix is known or can be estimated reliably, the dependence can be utilized to enhance the signal strength of the testing problem so as to improve the performance of the HC test. The improvement is reflected in lowering the needed signal strength r by a constant factor. Delaigle and Hall (2009) evaluated the HC test under a nonparametric setting allowing column-wise dependence, and showed that the detection boundary of Donoho and Jin (2004) for the HC test can be maintained under weak column-wise dependence. Delaigle, Hall and Jin (2011) showed that the standard HC test based on the normality assumption can perform poorly when the underlying data deviate from the normal distribution and studied a version of the HC test based on the t-statistics formulation. Cai, Jeng and Jin (2011) considered detecting Gaussian mixtures which differ from the null in both the mean and the variance. Arias-Castro, Bubeck and Lugosi (2012a, 2012b) established the lower and upper bounds for the minimax risk for detecting sparse differences in the covariance.

We show in this paper that there are alternative test procedures for weakly dependent sub-Gaussian data with unknown covariance which attain the same detection boundary as the HC test established in Donoho and Jin (2004) for Gaussian distributed data with $\boldsymbol{\Sigma} = \mathbf{I}_p$. The alternative test statistics are obtained by first constructing, for $\gamma = 1$ and 2,

$$T_{\gamma n}(s) = \sum_{j=1}^{p} |\sqrt{n}\bar{X}_j/\sigma_j|^{\gamma} I\bigl(|\bar{X}_j| \geq \sigma_j\sqrt{\lambda_p(s)/n}\bigr),$$

which threshold with respect to $\bar{X}_j$ at a level $\sqrt{\lambda_p(s)/n}$ for $s \in (0,1)$, where $\lambda_p(s) = 2s \log p$, $\bar{X}_j$ is the sample mean of the jth margin of the data vectors and $I(\cdot)$ is the indicator function. We note that $\gamma = 1$ and 2 correspond to the L_1 and L_2 versions of the thresholding statistics, respectively; and $\gamma = 0$ corresponds to the HC test statistic. In the literature, the L_1 statistic is called the hard thresholding in Donoho and Johnstone (1994) and Donoho and Jin (2008), and the L_0 statistic is called the clipping thresholding in Donoho and Jin (2008). We then maximize standardized versions of $T_{\gamma n}(s)$ with respect to s over $\mathcal{S}$, a subset of (0, 1), which results in the following maximal L_γ-thresholding statistics:

$$\hat{\mathcal{M}}_{\gamma n} = \max_{s \in \mathcal{S}} \frac{T_{\gamma n}(s) - \hat{\mu}_{T_{\gamma n},0}(s)}{\hat{\sigma}_{T_{\gamma n},0}(s)} \qquad \text{for } \gamma = 0, 1 \text{ and } 2, \tag{1.3}$$

where $\hat{\mu}_{T_{\gamma n},0}(s)$ and $\hat{\sigma}_{T_{\gamma n},0}(s)$ are, respectively, estimators of the mean $\mu_{T_{\gamma n},0}(s)$ and standard deviation $\sigma_{T_{\gamma n},0}(s)$ of $T_{\gamma n}(s)$ under H_0, whose forms will be given later in the paper. By developing the asymptotic distributions of $\hat{\mathcal{M}}_{\gamma n}$, the maximal L_γ-thresholding tests are formulated for $\gamma = 0, 1$ and 2 with the maximal L_0-test being equivalent to the HC test. An analysis on the relative power performance of the three tests reveals that if the signal strength parameter $r \in (0,1)$, the maximal L_2-thresholding test is at least as powerful as the maximal L_1-thresholding test, and both the L_1 and L_2-thresholding tests are at least as powerful as the HC test. If we allow a slightly stronger signal so that $r > 2\beta - 1$, the differential power performance of the three tests is amplified with the maximal L_2-test being the most advantageous followed by the maximal L_1-test.

In addition to the connection to the HC test, the maximal L_γ-thresholding test, by its nature of formulation, is related to the high-dimensional multivariate testing procedures, for instance, the tests proposed by Bai and Saranadasa (1996) and Chen and Qin (2010). While these tests can maintain accurate size approximation under a diverse range of dimensionality and column-wise dependence, their performance is hampered when the nonzero means are sparse and faint. The proposed test formulation is also motivated by a set of earlier works including Donoho and Johnstone (1994) for selecting significant wavelet coefficients, and Fan (1996) who considered testing for the mean of a random vector $\mathbf{X}$ with I.I.D. normally distributed components. We note that the second step of maximization with respect to $s \in \mathcal{S} \subset (0,1)$ is designed to make the test adaptive to the underlying signals strength and sparsity, which is the essence of the HC procedure in Donoho and Jin (2004), as well as that of Fan (1996).

The rest of the paper is organized as follows. In Section 2 we provide basic results on the L_2-thresholding statistic via the large deviation method and the asymptotic distribution of the single threshold statistic. Section 3 gives the asymptotic distribution of $\hat{\mathcal{M}}_{2n}$ as well as the associated test procedure. Power comparisons among the HC and the maximal L_1 and L_2-thresholding tests are made in Section 4. Section 5 reports simulation results which confirm the theoretical results.

Some discussions are given in Section 6. All technical details are relegated to the Appendix.

2. Single threshold test statistic. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an independent p-dimensional random sample from a common distribution F, and $\mathbf{X}_i = \mathbf{W}_i + \boldsymbol{\mu}$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$ is the vector of means and $\mathbf{W}_i = (W_{i1}, \ldots, W_{ip})^T$ is a vector consisting of potentially dependent random variables with zero mean and finite variances. The dependence among $\{W_{ij}\}_{j=1}^p$ is called the column-wise dependence in $\mathbf{W}_i$. Those nonzero μ_j are called "signals."

Let $\bar{X}_j = n^{-1}\sum_{i=1}^n X_{ij}$, $\sigma_j^2 = \operatorname{Var}(W_{ij})$ and $s_j^2 = (n-1)^{-1}\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2$ be the sample variance for the jth margin. The signal strength in the jth margin can be measured by the t-statistics $\sqrt{n}\bar{X}_j/s_j$ or the z-statistics $\sqrt{n}\bar{X}_j/\sigma_j$ if σ_j is known. For easy expedition, the test statistics will be constructed based on the z-statistics by assuming σ_j is known and, without loss of generality, we assume $\sigma_j^2 = 1$. Using the t-statistics actually leads to less restrictive conditions for the underlying random variables since the large deviation results for the self-normalized t-statistics can be established under weaker conditions to allow heavier tails in the underlying distribution as demonstrated in Shao (1997), Jing, Shao and Zhou (2008) and Wang and Hall (2009). See Delaigle, Hall and Jin (2011) for analysis on the sparse signal detection using the t-statistics.

We assume the following assumptions in our analysis:

(C.1) The dimension $p = p(n) \to \infty$ as $n \to \infty$ and $\log(p) = o(n^{1/3})$.

(C.2) There exists a positive constant H such that, for any $j \neq l \in \{1, \ldots, p\}$, $E(e^{h^T(W_{1j}^d, W_{1l}^d)}) < \infty$ for $h \in [-H, H] \times [-H, H]$ and $d = 2$.

(C.3) For each $i = 1, \ldots, n$, $\{W_{ij}\}_{j=1}^p$ is a weakly stationary sequence such that $E(W_{ij}) = E(W_{i(j+k)}) = 0$ and $\operatorname{Cov}(W_{ij}, W_{i(j+k)})$ does not depend on j for any integer k. And $\sum_k |\rho_k| < \infty$ where $\rho_k = \operatorname{Cov}(W_{i1}, W_{i(k+1)})$.

(C.4) Among the p marginal means, there are $m = p^{1-\beta}$ signals for a $\beta \in (1/2, 1)$ and the signal $\mu_j = \sqrt{2r\log(p)/n}$ for a $r > 0$. The signals' locations $\ell_1 < \ell_2 < \cdots < \ell_m$ are randomly selected from $\{1, 2, \ldots, p\}$ without replacement so that

$$P(\ell_1 = p_1, \ldots, \ell_m = p_m) = \binom{p}{m}^{-1} \quad \text{for all } 1 \le p_1 < p_2 < \cdots < p_m \le p. \tag{2.1}$$

(C.1) specifies the growth rate of p relative to the sample size n is in the paradigm of "large p, small n." That $\log p = o(n^{1/3})$ is the rate we can attain for Gaussian data or cases where we can attain "accurate" enough estimation of $\mu_{T_{\gamma n},0}$, which satisfies equation (2.6). When data are not Gaussian and the "accurate" estimators are not attainable, the growth rate of p will be more restrictive at $p = n^{1/\theta}$ ($\theta > 0$), as will be discussed in the next section. (C.2) assumes the joint distributions of (W_{ij}, W_{il}) is sub-Gaussian, which implies each marginal W_{ij} is sub-Gaussian as well. (C.3) prescribes weak dependence among $\{W_{ij}\}_{j=1}^p$.

The first part of (C.4) reiterates the sparse and faint signal setting. The range of the signal strength includes the case of $r \in (0, 1)$, representing the most fainted detectable signal strength, which has been considered in Donoho and Jin (2004) and other research works. The second part of (C.4) provides a random allocation mechanism for the signal bearing dimensions, which is the same as the one assumed in Hall and Jin (2010). Existing research on the detection boundary of the HC test for the sparse mean problem [Donoho and Jin (2004); Hall and Jin (2010)] is largely conducted for the case of $n = 1$ when the data are Gaussian. This is understandable since the sample means are sufficient statistics and there is no loss of generality when we treat the problem as $n = 1$, even if we have multiple observations. However, when the underlying distributions are as specified in (C.2), we cannot translate the test problem to $n = 1$ without incurring a loss of information.

We first consider the L_2 version of the thresholding statistic T_{2n} in this section. The study of the T_{1n} version is outlined in Section 4 when we compare the power performance to the HC test. Let $Y_{j,n} = n\bar{X}_j^2$. Then, the L_2-thresholding statistic can be written as

$$T_{2n}(s) = \sum_{j=1}^{p} Y_{j,n} I\{Y_{j,n} \geq \lambda_p(s)\}, \tag{2.2}$$

where s is the thresholding parameter that takes values over a range within (0, 1). There is no need to consider $s \geq 1$ in the thresholding since large deviation results given in Petrov (1995) imply that under H_0, $P(\max_{1\leq j\leq p} Y_{j,n} \leq \lambda_p(s)) \to 1$.

Define a set of slowing varying functions: $L_p^{(1)} = 2r\log p + 1$, $L_p^{(2)} = 2\sqrt{s\log p/\pi}$, $L_p^{(3)} = s(\sqrt{s} - \sqrt{r})^{-1}\sqrt{\log p/\pi}$, $L_p^{(4)} = 8r\log p$, $L_p^{(5)} = 4s^{3/2} \times \pi^{-1/2}(\log p)^{3/2}$ and $L_p^{(6)} = 2s^2(\log p)^{3/2}/\sqrt{\pi}(\sqrt{s} - \sqrt{r})$. Let $\phi(\cdot)$ and $\bar{\Phi}(\cdot)$ be the density and survival functions of the standard normal distribution.

Let $\mu_{T_{2n},0}(s)$ and $\sigma^2_{T_{2n},0}(s)$ be the mean and variance of $T_{2n}(s)$ under H_0, respectively, and $\mu_{T_{2n},1}(s)$ and $\sigma^2_{T_{2n},1}(s)$ be those, respectively, under the H_1 as specified in (C.4). The following proposition depicts the mean and variance of $T_{2n}(s)$ by applying Fubini's theorem and the large deviation results [Petrov (1995) and Lemma A.1 in Zhong, Chen and Xu (2013)].

PROPOSITION 1. *Under* (C.1)–(C.4), $E\{T_{2n}(s)\}$ *and* $\operatorname{Var}\{T_{2n}(s)\}$ *are, respectively,*

$$\begin{aligned}&\mu_{T_{2n},0}(s)\\&\quad = p\{2\lambda_p^{1/2}(s)\phi(\lambda_p^{1/2}(s)) + 2\bar{\Phi}(\lambda_p^{1/2}(s))\}\{1 + O\{n^{-1/2}\lambda_p^{3/2}(s)\}\},\end{aligned} \tag{2.3}$$

$$\begin{aligned}&\sigma^2_{T_{2n},0}(s)\\&\quad = p\{2[\lambda_p^{3/2}(s) + 3\lambda_p^{1/2}(s)]\phi(\lambda_p^{1/2}(s)) + 6\bar{\Phi}(\lambda_p^{1/2}(s))\}\{1 + o(1)\}\end{aligned} \tag{2.4}$$

under the H_0; *and*

$$\mu_{T_{2n},1}(s) = \{L_p^{(1)} p^{1-\beta} I(s < r) + L_p^{(3)} p^{1-\beta-(\sqrt{s}-\sqrt{r})^2} I(s > r)\}\{1 + o(1)\} + \mu_{T_{2n},0}(s),$$

$$\sigma^2_{T_{2n},1}(s) = \{L_p^{(4)} p^{1-\beta} I(s < r) + L_p^{(5)} p^{1-s} + L_p^{(6)} p^{1-\beta-(\sqrt{s}-\sqrt{r})^2} I(s > r)\} \times \{1 + o(1)\}$$

under the H_1 *specified in* (C.4).

Expressions (2.3) and (2.4) provide the first and the second order terms of $\mu_{T_{2n},0}(s)$ and $\sigma^2_{T_{2n},0}(s)$, which are needed when we consider their empirical estimation under H_0 when formulating the L_2 thresholding test statistic. Note that $\mu_{T_{2n},0}(s) = L_p^{(2)} p^{1-s}\{1 + o(1)\}$ and $\sigma^2_{T_{2n},0}(s) = L_p^{(5)} p^{1-s}\{1 + o(1)\}$. Only the first order terms for the variance are needed under H_1, but the approximation to $\mu_{T_{2n},1}(s)$ has to be more accurate so as to know the order of the difference between $\mu_{T_{2n},1}(s)$ and $\mu_{T_{2n},0}(s)$. Proposition 1 indicates that the column-wise dependence as specified in (C.3) does not have much leading order impact on the variance of $T_{2n}(s)$. The leading order variance is almost the same when $\mathbf{W}_i$ are column-wise independent. The difference only appears in the coefficients of the slow-varying functions $L_p^{(4)}$, $L_p^{(5)}$ and $L_p^{(6)}$, while their orders of magnitude remain unchanged. The reason behind this phenomena is the thresholding. It can be understood by an analogue for multivariate Gaussian distributions with nonzero correlation. Despite the dependence in the Gaussian distribution, exceedances beyond high thresholds are asymptotically independent [Sibuya (1960) and Joe (1997)].

We now study the asymptotic distribution of $T_{2n}(s)$ to prepare for the proposal of the maximal L_2-thresholding statistic. Write

$$T_{2n}(s) = \sum_{j=1}^{p} Z_{j,n}(s),$$

where $Z_{j,n}(s) := Y_{j,n} I\{Y_{j,n} > \lambda_p(s)\}$ and $\lambda_p(s) = 2s \log(p)$. For integers $a, b \in [-\infty, \infty]$ such that $a < b$, define $\mathscr{F}_a^b = \sigma\{Z_{l,n}(s) : l \in (a, b)\}$ as the σ-algebra generated by $\{Z_{l,n}(s)\}_{l=a}^b$ and define the ρ-mixing coefficients

$$\rho_{Z(s)}(k) = \sup_{l, \xi \in L^2(\mathscr{F}_{-\infty}^l), \zeta \in L^2(\mathscr{F}_{l+k}^\infty)} |\mathrm{Corr}(\xi, \zeta)|. \tag{2.5}$$

See Doukhan (1994) for comprehensive discussions on the mixing concept. The following is a condition regarding the dependence among $\{Z_{j,n}(s)\}_{j=1}^p$.

(C.5) For any $s \in (0, 1)$, the sequence of random variables $\{Z_{j,n}(s)\}_{j=1}^p$ is ρ-mixing such that $\rho_{Z(s)}(k) \le C\alpha^k$ for some $\alpha \in (0, 1)$ and a positive constant C.

The requirement of $\{Z_{j,n}(s)\}_{j=1}^{p}$ being ρ-mixing for each s is weaker than requiring the original data columns $\{X_{ij}\}_{j=1}^{p}$ being ρ-mixing, whose mixing coefficient $\rho_{X_i}(k)$ can be similarly defined as (2.5). This is because, according to Theorem 5.2 in Bradley (2005),

$$\rho_{Z(s)}(k) \le \sup_{i\le n} \rho_{X_i}(k) = \rho_{X_1}(k) \qquad \text{for each } k = 1, \dots, p \text{ and } s \in (0, 1).$$

The following theorem reports the asymptotic normality of $T_{2n}(s)$ under both H_0 and H_1.

THEOREM 1. *Assume* (C.1)–(C.5). *Then, for any* $s \in (0, 1)$,

$$\text{(i)}\quad \sigma_{T_{2n},0}^{-1}(s)\{T_{2n}(s) - \mu_{T_{2n},0}(s)\} \xrightarrow{d} N(0, 1) \qquad \textit{under } H_0;$$

$$\text{(ii)}\quad \sigma_{T_{2n},1}^{-1}(s)\{T_{2n}(s) - \mu_{T_{2n},1}(s)\} \xrightarrow{d} N(0, 1) \qquad \textit{under } H_1.$$

From (2.3) and (2.4), define the leading order terms of $\mu_{T_{2n},0}(s)$ and $\sigma^2_{T_{2n},0}(s)$, respectively,

$$\tilde\mu_{T_{2n},0}(s) = p\{2\lambda_p^{1/2}(s)\phi(\lambda_p^{1/2}(s)) + 2\bar\Phi(\lambda_p^{1/2}(s))\} \qquad \text{and}$$
$$\tilde\sigma^2_{T_{2n},0}(s) = p\{2[\lambda_p^{3/2}(s) + 3\lambda_p^{1/2}(s)]\phi(\lambda_p^{1/2}(s)) + 6\bar\Phi(\lambda_p^{1/2}(s))\}.$$

It is clear that the asymptotic normality in Theorem 1(i) remains if we replace $\sigma_{T_{2n},0}(s)$ by $\tilde\sigma_{T_{2n},0}(s)$.

To formulate a test procedure based on the thresholding statistic $T_{2n}(s)$, we need to estimate $\mu_{T_{2n},0}(s)$ by a $\hat\mu_{T_{2n},0}(s)$, say. Ideally, if

$$\mu_{T_{2n},0}(s) - \hat\mu_{T_{2n},0}(s) = o\{\tilde\sigma_{T_{2n},0}(s)\}, \tag{2.6}$$

the first part of Theorem 1 remains valid if we replace $\mu_{T_{2n},0}(s)$ with $\hat\mu_{T_{2n},0}(s)$. An obvious choice of $\hat\mu_{T_{2n},0}(s)$ is $\tilde\mu_{T_{2n},0}(s)$, which is known upon given p and s. Indeed, if W_{ij}s are the standard normally distributed, we have

$$\mu_{T_{2n},0}(s) = \tilde\mu_{T_{2n},0}(s) \qquad \text{for } s \in (0, 1),$$

implying the leading order is exactly $\mu_{T_{2n},0}(s)$ for the Gaussian data. Hence, if we take $\hat\mu_{T_{2n},0}(s) = \tilde\mu_{T_{2n},0}(s)$, (2.6) is satisfied for the Gaussian data.

For non-Gaussian observations, the difference between $\mu_{T_{2n},0}(s)$ and $\tilde\mu_{T_{2n},0}(s)$ may not be a smaller order of $\sigma_{T_{2n},0}(s)$. Specifically, from (2.3) and (2.4), we have

$$\frac{\mu_{T_{2n},0}(s) - \tilde\mu_{T_{2n},0}(s)}{\sigma_{T_{2n},0}(s)} = O\{\lambda_p^{5/4}(s) p^{(1-s)/2} n^{-1/2}\}.$$

To make the above ratio diminishing to zero, the strategy of Delaigle, Hall and Jin (2011) can be adopted by restricting $p = n^{1/\theta}$ and $s \in ((1-\theta)_+, 1)$ for a positive θ, where $(a)_+ = a$ if $a > 0$ and $(a)_+ = 0$ if $a \le 0$. Under this circumstance,

$$\frac{\mu_{T_{2n},0}(s) - \tilde\mu_{T_{2n},0}(s)}{\sigma_{T_{2n},0}(s)} = O\{(2s/\theta \log n)^{5/4} n^{(1-s-\theta)/(2\theta)}\} \to 0. \tag{2.7}$$

Clearly, for a not so high dimension with $\theta \geq 1$, (2.7) holds for all $s \in (0, 1)$, and $\tilde{\mu}_{T_{2n},0}(s)$ satisfies (2.6). For higher dimensions with $\theta < 1$, the thresholding level s has to be restricted to ensure (2.7). The restriction can alter the detection boundary of the test we will propose in the next section. This echoes a similar phenomena for the HC test given in Delaigle, Hall and Jin (2011). To expedite our discussion, we assume in the rest of the paper that (2.6) is satisfied by the $\hat{\mu}_{T_{2n},0}(s)$. We note such an arrangement is not entirely unrealistic, as a separate effort may be made to produce more accurate estimators. Assuming so allows us to stay focused on the main agenda of the testing problem.

The asymptotic normality established in Theorem 1 allows an asymptotic α-level test that rejects H_0 if

$$T_{2n}(s) - \hat{\mu}_{T_{2n},0}(s) > z_\alpha \tilde{\sigma}_{T_{2n},0}(s), \tag{2.8}$$

where z_α is the upper α quantile of the standard normal distribution.

3. Maximal thresholding. While the asymptotic normality of $T_{2n}(s)$ in Theorem 1 ensures the single thresholding level test in (2.8) a correct size asymptotically, the power of the test depends on s, the underlying signal strength r and the sparsity β. A test procedure is said to be able to separate a pair of null and alternative hypotheses asymptotically if the sum of the probabilities of the type I and type II errors converges to zero as $n \to \infty$. Let α_n be a sequence of the probabilities of type I error, which can be made converging to zero as $n \to \infty$. The sum of the probabilities of the type I and type II errors for the test given in (2.8) with nominal size α_n is approximately

$$\mathrm{Err}_{\alpha_n} := \alpha_n + P\left(\frac{T_{2n}(s) - \mu_{T_{2n},0}(s)}{\sigma_{T_{2n},0}(s)} \leq z_{\alpha_n} \middle| H_1\right), \tag{3.1}$$

which is attained based on the facts that (i) the size α_n is attained asymptotically and (ii) $\hat{\mu}_{T_{2n},0}(s)$ and $\tilde{\sigma}_{T_{2n},0}(s)$ are sufficiently accurate estimators in the test procedure (2.8).

Our strategy is to first make $\alpha_n \to 0$ such that $z_{\alpha_n} = C(\log p)^\varepsilon$ for an arbitrarily small $\varepsilon > 0$ and a constant $C > 0$. The second term on the right-hand side of (3.1) is

$$\begin{aligned} \mathrm{Err}_{II} := P\biggl(&\frac{T_{2n}(s) - \mu_{T_{2n},1}(s)}{\sigma_{T_{2n},1}(s)} \\ &\leq z_{\alpha_n} \frac{\sigma_{T_{2n},0}(s)}{\sigma_{T_{2n},1}(s)} - \frac{\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s)}{\sigma_{T_{2n},1}(s)}\biggr). \end{aligned} \tag{3.2}$$

Because z_{α_n} is slowly varying, $0 < \sigma_{T_{2n},0}(s)/\sigma_{T_{2n},1}(s) \leq 1$ and $(T_{2n}(s) - \mu_{T_{2n},1}(s))/\sigma_{T_{2n},1}(s)$ is stochastically bounded, a necessary and sufficient condition that ensures $\mathrm{Err}_{\alpha_n} \to 0$ is

$$\Delta_2(s; r, \beta) := \frac{\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s)}{\sigma_{T_{2n},1}(s)} \to \infty. \tag{3.3}$$

From Proposition 1, it follows that, up to a factor $1 + o(1)$,

$$\Delta_2(s; r, \beta) = \begin{cases} C_1 p^{(1+s-2\beta)/2}, & \text{if } s \le r \text{ and } s \le \beta; \\ C_2 p^{(1-\beta)/2}, & \text{if } s \le r \text{ and } s > \beta; \\ C_3 p^{1/2-\beta+r-(\sqrt{s}-2\sqrt{r})^2/2}, & \text{if } s > r \text{ and } s \le (\sqrt{s}-\sqrt{r})^2 + \beta; \\ C_4 p^{(1-\beta-(\sqrt{s}-2\sqrt{r})^2)/2}, & \text{if } s > r \text{ and } s > (\sqrt{s}-\sqrt{r})^2 + \beta, \end{cases}$$

where $C_1 = \sqrt{2}(\pi s)^{1/4}(\frac{r}{s})(\log p)^{1/4}$, $C_2 = \frac{1}{2}(r \log p)^{1/2}$, $C_3 = s^{1/4}(\log p)^{-1/4}/\{\sqrt{2}\pi^{1/4}(\sqrt{s} - \sqrt{r})\}$ and $C_4 = (2\sqrt{\pi}(\sqrt{s} - \sqrt{r}))^{-1/2}(\log p)^{-1/4}$.

Let

$$\varrho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \le 3/4; \\ (1 - \sqrt{1-\beta})^2, & 3/4 < \beta < 1. \end{cases}$$

As demonstrated in Donoho and Jin (2004) and Ingster (1997), the phase diagram $r = \varrho^*(\beta)$ is the optimal detection boundary for testing the hypotheses we are considering in this paper when the data are Gaussian and $\mathbf{\Sigma} = \mathbf{I}_p$. Here the optimality means that for any $r > \varrho^*(\beta)$, there exists at least one test such that the sum of the probabilities of the type I and type II errors diminishes to zero as $n \to \infty$; but for $r < \varrho^*(\beta)$, no such test exists. For correlated Gaussian data such that $\mathbf{\Sigma} \ne \mathbf{I}_p$, Hall and Jin (2010) found that the detection boundary $r = \varrho^*(\beta)$ may be lowered by transforming the data via the inverse of Cholesky factorization $\mathbf{L}$ such that $\mathbf{L}\Sigma\mathbf{L}^T = \mathbf{I}_p$. More discussion on the optimality is given in Section 6.

From the expression of $\Delta_2(s; r, \beta)$ given above, it can be shown (see the proof of Theorem 3 in the Appendix) that if $r > \varrho^*(\beta)$ there exists at least one $s \in (0, 1)$ for each pair of (r, β) such that (3.3) is satisfied and, hence, the thresholding test would be powerful. This is the key for the maximal L_2-thresholding test that we will propose later to attain the detection boundary.

It is clear that we have to make the thresholding level s adaptive to the unknown r and β. One strategy is to use a range of thresholding levels, say, $s \in \mathcal{S} \subset (0, 1)$, so that the underlying (r, β) can be "covered." This is the very idea of the HC test.

Let $\hat{\mathcal{T}}_{2,n}(s) = \tilde{\sigma}^{-1}_{T_{2n},0}(s)\{T_{2n}(s) - \hat{\mu}_{T_{2n},0}(s)\}$ be the standardized version of $T_{2n}(s)$. Define the maximal thresholding statistic

$$\hat{\mathcal{M}}_{2n} = \sup_{s \in \mathcal{S}} \hat{\mathcal{T}}_{2,n}(s),$$

where $\mathcal{S} = (0, 1 - \eta]$ for an arbitrarily small positive η. Let

$$\mathcal{S}_n = \{s_i : s_i = Y_{i,n}/(2 \log p) \text{ and } 0 < Y_{i,n} < 2(1 - \eta) \log p\} \cup \{1 - \eta\}. \tag{3.4}$$

Since both $\hat{\mu}_{T_{2n},0}(s)$ and $\tilde{\sigma}_{T_{2n},0}(s)$ are monotone decreasing functions of s, it can be shown that $\hat{\mathcal{M}}_{2n}$ can be attained on $\mathcal{S}_n$, namely,

$$\hat{\mathcal{M}}_{2n} = \max_{s \in \mathcal{S}_n} \hat{\mathcal{T}}_{2,n}(s). \tag{3.5}$$

This largely reduces the computational burden of $\hat{\mathcal{M}}_{2n}$. The asymptotic distribution of $\hat{\mathcal{M}}_{2n}$ is established in the following theorem.

THEOREM 2. *Assume* (C.1)–(C.3), (C.5) *and* (2.6) *hold*. *Then*, *under* H_0,

$$P\big(a(\log p)\hat{\mathcal{M}}_{2n} - b(\log p, \eta) \le x\big) \to \exp(-e^{-x}),$$

where $a(y) = (2\log(y))^{1/2}$ *and* $b(y,\eta) = 2\log(y) + 2^{-1}\log\log(y) - 2^{-1} \times \log(\frac{4\pi}{(1-\eta)^2})$.

The theorem leads to an asymptotic α-level test that rejects H_0 if

$$\hat{\mathcal{M}}_{2n} > \mathcal{B}_\alpha = \big(\mathcal{E}_\alpha + b(\log p, \eta)\big)/a(\log p), \tag{3.6}$$

where $\mathcal{E}_\alpha$ is the upper α quantile of the Gumbel distribution $\exp(-e^{-x})$. We name the test the maximal L_2-thresholding test. The following theorem shows that its detection boundary is $r = \varrho^*(\beta)$.

THEOREM 3. *Under conditions* (C.1)–(C.5) *and assuming* (2.6) *holds*, *then* (i) *if* $r > \varrho^*(\beta)$, *the sum of the type I and II errors of the maximal* L_2*-thresholding tests converges to* 0 *when the nominal sizes* $\alpha_n = \bar{\Phi}((\log p)^{\varepsilon}) \to 0$ *for an arbitrarily small* $\varepsilon > 0$ *as* $n \to \infty$.

(ii) *If* $r < \varrho^*(\beta)$, *the sum of the type I and II errors of the maximal* L_2*-thresholding test converges to* 1 *when the nominal sizes* $\alpha_n \to 0$ *as* $n \to \infty$.

It is noted that when $r > \varrho^*(\beta)$ in part (i) of Theorem 3, we need to restrict the rate of the nominal type I error α_n's convergence to 0, since the conclusion of part (i) may not be true for all $\alpha_n \to 0$. However, in part (ii) where $r < \varrho^*(\beta)$, no restriction for α_n is required, which has to be the case, as otherwise there is no guarantee that $r = \varrho^*(\beta)$ is the detection boundary of the test.

If the estimator $\hat{\mu}_{T_{2n,0}}(s)$ cannot attain (2.6) and $\tilde{\mu}_{T_{2n,0}}(s)$ is used as the estimator, we have to restrict $p = n^{1/\theta}$ for a $\theta \in (0,1)$ and limit $s \in (1-\theta, 1)$. In this case, the above theorem is valid if we replace $\varrho^*(\beta)$ by $\varrho^*_\theta(\beta)$, where

$$\varrho^*_\theta(\beta) = \begin{cases} (\sqrt{1-\theta} - \sqrt{1-\beta-\theta/2})^2, & \text{if } 1/2 < \beta \le (3-\theta)/4; \\ \beta - 1/2, & \text{if } (3-\theta)/4 < \beta \le 3/4; \\ (1-\sqrt{1-\beta})^2, & \text{if } 3/4 < \beta < 1, \end{cases}$$

which is clearly inferior to $\varrho^*(\beta)$. The boundary $\varrho^*_\theta(\beta)$ is the same as the one in Delaigle, Hall and Jin (2011) based on the marginal t-statistics, whereas our result is based on the z-statistics. The t-statistic formulation reduces the demand on the tails of the distributions as shown in Delaigle, Hall and Jin (2011). We note that if $\theta \ge 1$, Theorem 3 remains so that the Gaussian detection boundary is still valid.

4. Power comparison. We compare the power of the maximal L_2-thresholding test with those of the HC test and the maximal L_1-thresholding test in this section. Let us first introduce these two tests.

The HC test is based on

$$\hat{\mathcal{T}}_{0,n}(s) = \frac{T_{0n}(s) - 2p\bar{\Phi}(\lambda_p^{1/2}(s))}{\sqrt{2p\bar{\Phi}(\lambda_p^{1/2}(s))(1 - 2\bar{\Phi}(\lambda_p^{1/2}(s)))}}, \tag{4.1}$$

where $T_{0n}(s) = \sum_{j=1}^{p} I(Y_{j,n} \geq \lambda_p(s))$. Like Delaigle and Hall (2009), we consider here a two-sided HC test instead of a one-sided test treated in Donoho and Jin (2004). With the same reasoning as Donoho and Jin [(2004), page 968], we define the HC test statistic

$$\hat{\mathcal{M}}_{0n} = \max_{s \in \mathcal{S}} \hat{\mathcal{T}}_{0,n}(s),$$

where $\mathcal{S} = (0, 1 - \eta]$ for an arbitrary small η and is the same as the maximal L_2-thresholding statistic. Using the same argument for the maximal L_2-thresholding statistic, it can be shown that $\hat{\mathcal{M}}_{0n}$ attains its maximum value on $\mathcal{S}_n$ given in (3.4) as well.

According to Donoho and Jin (2004), under H_0,

$$P\big(a(\log p)\hat{\mathcal{M}}_{0n} - b(\log p, \eta) \leq x\big) \to \exp\big(-e^{-x}\big),$$

with the same normalizing sequences as those in Theorem 2. Let $\mathcal{B}_\alpha$ be the same as that of the maximal L_2-thresholding test given in (3.6). An α level HC test rejects H_0 if

$$\hat{\mathcal{M}}_{0n} > \mathcal{B}_\alpha. \tag{4.2}$$

Let us introduce the maximal L_1-thresholding test statistic. Recall that

$$T_{1n}(s) = \sum_{j=1}^{p} |\sqrt{n}\bar{X}_j| I\big(|\bar{X}_j| > \sqrt{\lambda_p(s)/n}\big).$$

It can be shown that the mean and variance of $T_{1n}(s)$ under H_0 are, respectively,

$$\mu_{T_{1n},0}(s) = \sqrt{2/\pi}\, p^{1-s}\{1 + o(1)\} \quad \text{and}$$
$$\sigma^2_{T_{1n},0}(s) = \big\{2p^{1-s}\sqrt{(s/\pi)\log p}\big\}\{1 + o(1)\}.$$

Define

$$\hat{\mathcal{T}}_{1,n}(s) = \frac{T_{1n}(s) - \hat{\mu}_{T_{1n},0}(s)}{\tilde{\sigma}_{T_{1n},0}(s)},$$

where $\hat{\mu}_{T_{1n},0}(s)$ is a sufficiently accurate estimator of $\mu_{T_{1n},0}(s)$ in a similar sense to (2.6) and $\tilde{\sigma}^2_{T_{1n},0}(s) = 2p^{1-s}\sqrt{(s/\pi)\log p}$. The maximal L_1-thresholding statistic is

$$\hat{\mathcal{M}}_{1n} = \max_{s \in \mathcal{S}} \hat{\mathcal{T}}_{1,n}(s),$$

where, again, $\mathcal{S} = (0, 1-\eta]$. It can be shown that $\hat{\mathcal{M}}_{1n} = \max_{s\in\mathcal{S}_n} \hat{\mathcal{T}}_{1,n}(s)$ for the same S_n in (3.4).

Using a similar approach to that in Theorem 2, we can show that

$$P\big(a(\log p)\hat{\mathcal{M}}_{1n} - b(\log p, \eta) \le x\big) \to \exp\big(-e^{-x}\big).$$

Hence, an α-level maximal L_1-thresholding test rejects the H_0 if

$$\hat{\mathcal{M}}_{1n} > \mathcal{B}_\alpha. \tag{4.3}$$

From (3.6), (4.2) and (4.3), the three tests have the same critical values $\mathcal{B}_\alpha$ at nominal level α. This brings convenience for the power comparison. Let us define the power of the three tests

$$\Omega_\gamma(r, \beta) := P(\hat{\mathcal{M}}_{\gamma n} > \mathcal{B}_\alpha)$$

for $\gamma = 0, 1$ and 2, respectively. Notice that

$$\hat{\mathcal{M}}_{\gamma n} = \max_{s\in\mathcal{S}_n}\big\{\mathcal{T}_{\gamma n}(s)\tilde{e}_\gamma(s) + \tilde{\sigma}^{-1}_{T_{\gamma n},0}(s)\big(\mu_{T_{\gamma n},0}(s) - \hat{\mu}_{T_{\gamma n},0}(s)\big)\big\}, \tag{4.4}$$

where $\tilde{e}_\gamma(s) = \sigma_{T_{\gamma n},0}(s)/\tilde{\sigma}_{T_{\gamma n},0}(s)$ and

$$\mathcal{T}_{\gamma n}(s) = \sigma^{-1}_{T_{\gamma n},0}(s)\big(T_{\gamma n}(s) - \mu_{T_{\gamma n},0}(s)\big) = \mathcal{T}_{\gamma n,1}(s)R_\gamma(s) + \Delta_{\gamma,0}(s; r, \beta),$$

in which $R_\gamma(s) = \sigma_{T_{\gamma n},1}(s)/\sigma_{T_{\gamma n},0}(s)$, $\mathcal{T}_{\gamma n,1}(s) = \sigma^{-1}_{T_{\gamma n},1}(s)(T_{\gamma n}(s) - \mu_{T_{\gamma n},1}(s))$ and $\Delta_{\gamma,0}(s; r, \beta) = \sigma^{-1}_{T_{\gamma n},0}(s)(\mu_{T_{\gamma n},1}(s) - \mu_{T_{\gamma n},0}(s))$. As shown in (A.8), (A.22) and (A.24) in the Appendix,

$$\begin{aligned}\Delta_{0,0}(s; r, \beta) &= (s\pi \log p)^{1/4} p^{1/2-\beta+s/2} I(r > s)\\ &\quad + L_p^{(6)} p^{1/2-\beta-(\sqrt{s}-\sqrt{r})^2+s/2} I(r < s),\\ \Delta_{1,0}(s; r, \beta) &= (s\pi \log p)^{1/4}(r/s)^{1/4} p^{1/2-\beta+s/2} I(r > s)\\ &\quad + L_p^{(6)} p^{1/2-\beta-(\sqrt{s}-\sqrt{r})^2+s/2} I(r < s)\end{aligned}$$

and

$$\begin{aligned}\Delta_{2,0}(s; r, \beta) &= (s\pi \log p)^{1/4}(r/s) p^{1/2-\beta+s/2} I(r > s)\\ &\quad + L_p^{(6)} p^{1/2-\beta-(\sqrt{s}-\sqrt{r})^2+s/2} I(r < s),\end{aligned}$$

where $L_p^{(6)} = \{2(\sqrt{s} - \sqrt{r})\}^{-1} s^{1/4}(\pi \log p)^{-1/4}$.

Derivations given in the proof of Theorem 4 in the Appendix show that for $\gamma = 0, 1$ and 2,

$$\hat{\mathcal{M}}_{\gamma n} \sim \max_{s\in\mathcal{S}_n} \Delta_{\gamma,0}(s; r, \beta), \tag{4.5}$$

where "$a \sim b$" means that the $a/b = 1 + o_p(1)$. This implies that we only need to compare $\max_{s\in\mathcal{S}_n} \Delta_{\gamma,0}(s; r, \beta)$ in the power comparison.

From the established expressions of $\Delta_{\gamma,0}(s;r,\beta)$, we note two facts. One is that if $r > 2\beta - 1$, for any $s \in (2\beta - 1, r)$,

$$\Delta_{2,0}(s;r,\beta)/\Delta_{1,0}(s;r,\beta) = (r/s)^{3/4} > 1 \quad \text{and} \tag{4.6}$$
$$\Delta_{1,0}(s;r,\beta)/\Delta_{0,0}(s;r,\beta) = (r/s)^{1/4} > 1.$$

The other is if $r \in (\varrho^*(\beta), 2\beta - 1]$, asymptotically,

$$\Delta_{0,0}(s;r,\beta) = \Delta_{1,0}(s;r,\beta) = \Delta_{2,0}(s;r,\beta) \qquad \text{for all } s \in \mathcal{S}. \tag{4.7}$$

Hence, when (r, β) lies just above the detection boundary, the three $\Delta_{\gamma,0}$ functions are the same. If (r, β) moves further away from the detection boundary so that $r > 2\beta - 1$, there will be a clear ordering among the $\Delta_{\gamma,0}$ functions. The following theorem summarizes the relative power performance.

THEOREM 4. *Assume* (C.1)–(C.5) *and* (2.6) *hold. For any given significant level* $\alpha \in (0, 1)$, *the powers of the HC, the maximal* L_1 *and* L_2*-thresholding tests under* H_1 *as specified in* (C.4) *satisfy, as* $n \to \infty$,

$$\Omega_0(r,\beta) \le \Omega_1(r,\beta) \le \Omega_2(r,\beta) \qquad \textit{for } r > 2\beta - 1 \tag{4.8}$$

and $\Omega_\gamma(r,\beta)$*s are asymptotic equivalent for* $r \in (\varrho^*(\beta), 2\beta - 1]$.

The theorem indicates that when (r, β) is well above the detection boundary such that $r > 2\beta - 1$, there is a clear ordering in the power among the three tests, with the L_2 being the most powerful followed by the L_1 test. However, when (r, β) is just above the detection boundary such that $r \in (\varrho^*(\beta), 2\beta - 1]$, the three tests have asymptotically equivalent powers. In the latter case, comparing the second order terms of $\hat{\mathcal{M}}_{\gamma n}$ may lead to differentiations among the powers of the three tests. However, it is a rather technical undertaking to assess the impacts of the second order terms. The analysis conducted in Theorem 4 is applicable to the setting of Gaussian data with $n = 1$ and $\mathbf{\Sigma}$ satisfying (C.3), which is the setting commonly assumed in the investigation of the detection boundary for the HC test [Donoho and Jin (2004); Hall and Jin (2010) and Arias-Castro, Bubeck and Lugosi (2012a)]. Specifically, the power ordering among the three maximal thresholding tests in Theorem 4 remains but under lesser conditions (C.3)–(C.5). Condition (C.1) is not needed since the Gaussian assumption allows us to translate the problem to $n = 1$ since the sample mean is sufficient. Condition (C.2) is automatically satisfied for the Gaussian distribution. The condition (2.6) is met for the Gaussian data, as we have discussed in Section 2.

5. Simulation results. We report results from simulation experiments which were designed to evaluate the performance of the maximal L_1 and L_2-thresholding tests and the HC test. The purpose of the simulation study is to confirm the theoretical findings that there is an ordering in the power among the three tests discovered in Theorem 4.

Independent and identically distributed p-dim random vectors $\mathbf{X}_i$ were generated according to

$$\mathbf{X}_i = \mathbf{W}_i + \boldsymbol{\mu}, \qquad i = 1, \dots, n,$$

where $\mathbf{W}_i = (W_{i1}, \dots, W_{ip})^T$ is a stationary random vector and $\{W_{ij}\}_{j=1}^{p}$ have the same marginal distribution F. In the simulation, $\mathbf{W}_i$ was generated from a p-dimensional multivariate Gaussian distribution with zero mean and covariance $\boldsymbol{\Sigma} = (\sigma_{ij})_{p\times p}$, where $\sigma_{ij} = \rho^{|i-j|}$ for $\rho = 0.3$ and 0.5, respectively.

The simulation design on $\boldsymbol{\mu}$ had the sparsity parameter $\beta = 0.6, 0.7$ and 0.8, respectively, and the signal strength $r = 0.1, 0.3, 0.5, 0.6, 0.8, 0.9, 1.1$ and 1.2, respectively. We chose two scenarios on the dimension and sample size combinations: (a) a large p, small n setting and (b) both p and n are moderately large. For scenario (a), we chose $p = \exp(c_0 n^{0.3} + c_1)$, where $c_0 = 1.90$ and $c_1 = 2.30$ so that the dimensions p were 2000 and 20,000, and the sample sizes n were 30 and 100, respectively. We note that under the setting $\beta = 0.8$, there were only 4 and 7 nonzero means, respectively, among the 2000 and 20,000 dimensions. And those for $\beta = 0.7$ were 9 and 19, respectively, and those for $\beta = 0.6$ were 20 and 52, respectively. These were quite sparse. For scenario (b), we chose $p = n^{1.25} + 184$ such that $(p, n) = (500, 100)$ and $(p, n) = (936, 200)$.

The maximal L_2-test statistic $\hat{\mathcal{M}}_{2n}$ was constructed using $\tilde{\mu}_{T_{2n},0}(s)$ and $\tilde{\sigma}_{T_{2n},0}(s)$ given in (2.3) and (2.4), respectively, as the mean and standard deviation estimators. The maximal L_1 test statistic and the HC test statistic, $\hat{\mathcal{M}}_{1n}$ and $\hat{\mathcal{M}}_{0n}$, were constructed similarly using the leading order mean and standard deviation under H_0. The set of thresholding level $\mathcal{S}$ was chosen to be $(0, 1-\eta]$ with $\eta = 0.05$.

Figures 1–4 display the average empirical sizes and powers of the HC, the maximal L_1 and L_2-thresholding tests based on 20,000 simulations, with Figures 1–2 for scenario (a) and Figures 3–4 for scenario (b). To make the power comparison fair and conclusive, we adjusted the nominal level of the tests so that the simulated sizes of the tests were all around $\alpha = 0.05$, with the HC having slightly larger sizes than those of the maximal L_1 test, and the sizes of the maximal L_1 test were slightly larger than those of the maximal L_2 test. These were designed to rule out potential "favoritism" in the power comparison due to advantages in the sizes of the maximal L_2 and/or L_1 tests.

Figures 1–4 show that the power of the tests were the most influenced by the signal strength parameter r, followed by the sparsity β. The powers were insensitive to the level of dependence ρ, which confirmed our finding that the thresholding largely removes the dependence. The observed ordering in the empirical power shown in Figures 1–4 were consistent to the conclusions in Theorem 4. We observed that in all the simulation settings, despite some size advantages by the HC test and/or the maximal L_1 test, the maximal L_2 test had better power than the maximal L_1 and the HC test, and the maximal L_1 test had better power than the

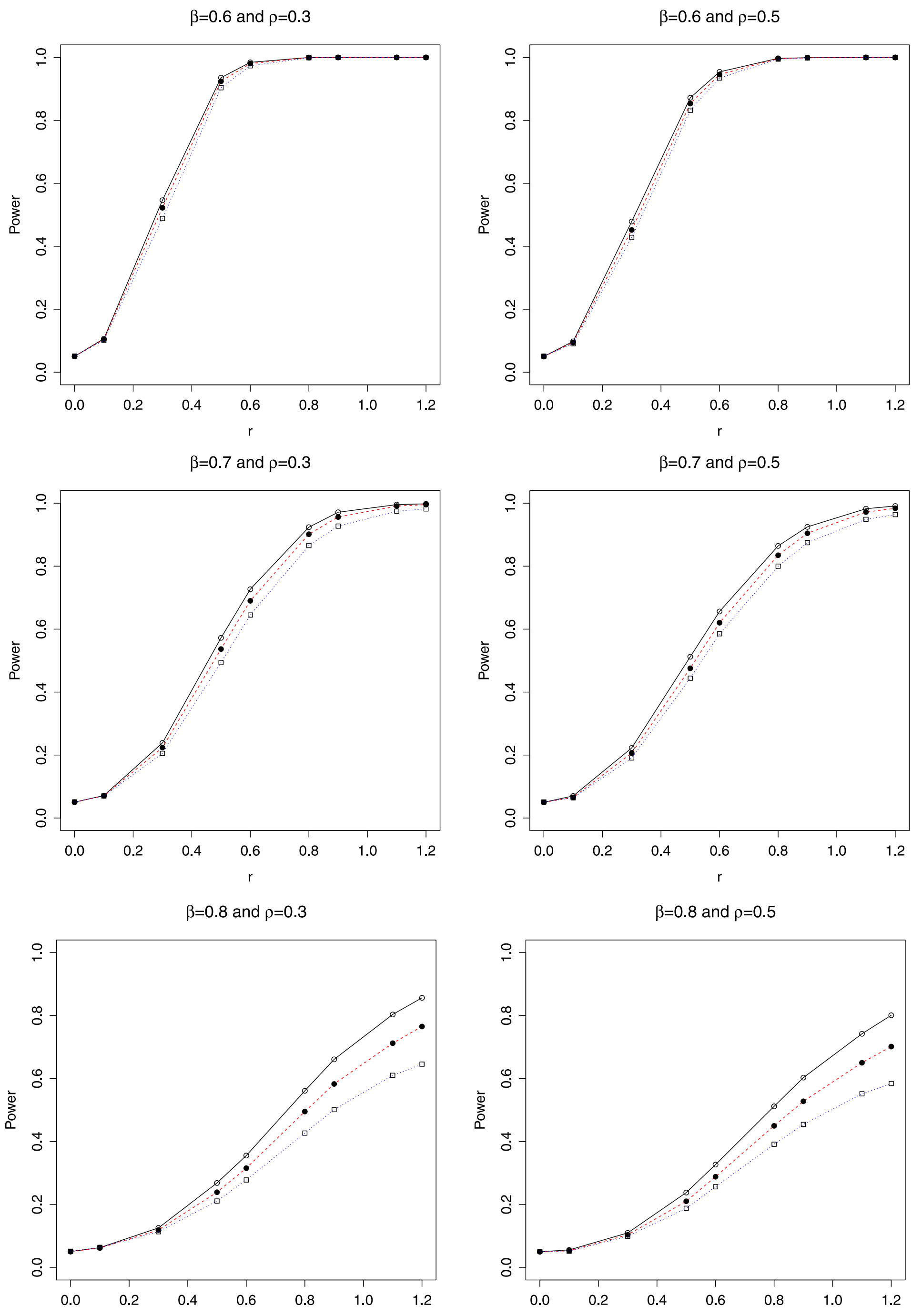

FIG. 1. *Empirical sizes and powers of the HC (dotted lines with squares), the maximal L_1- (dashed lines with dots) and L_2- (solid lines with circles) thresholding tests when $p = 2000$ and $n = 30$ with the marginal distribution the standard normal.*

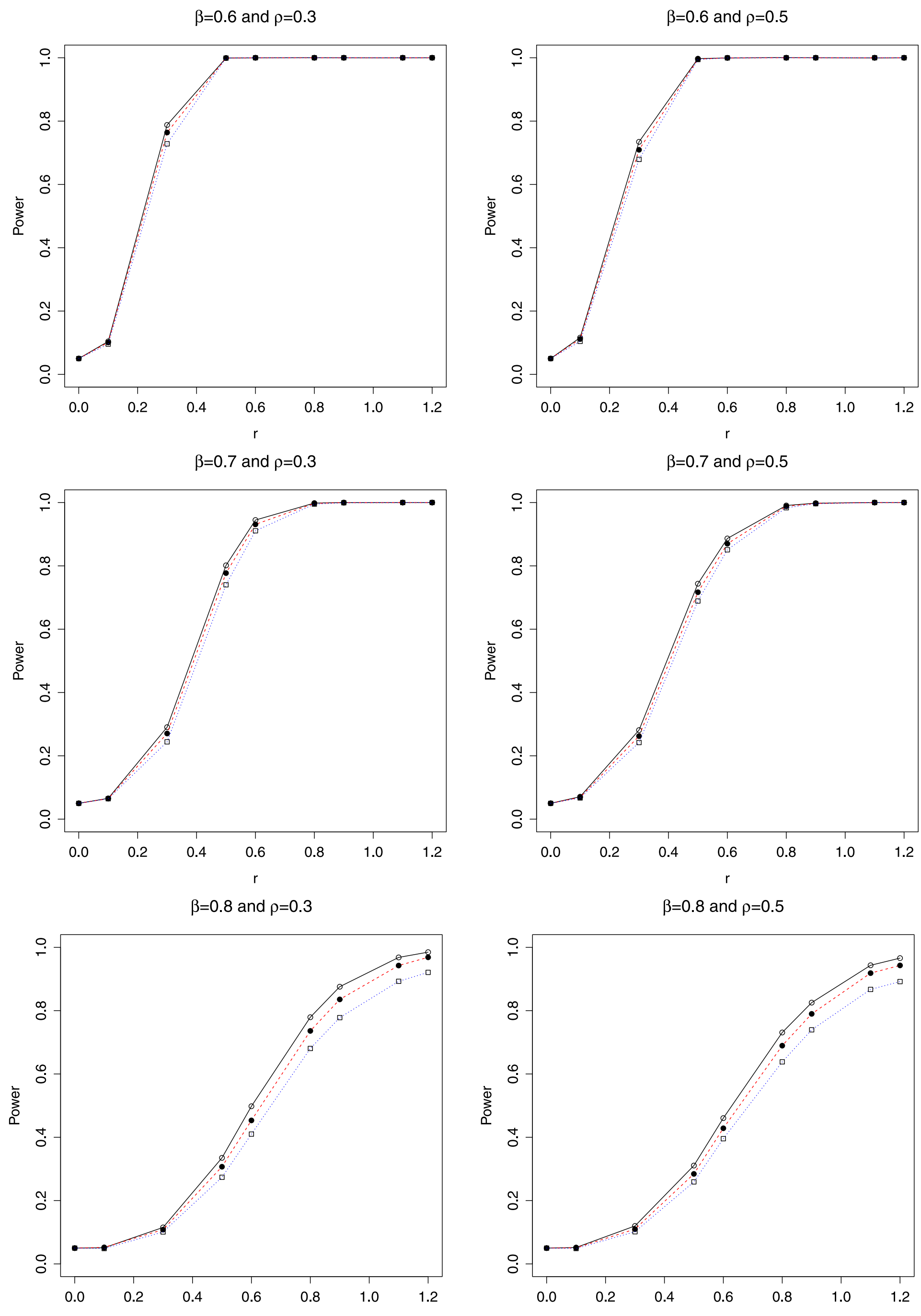

FIG. 2. *Empirical sizes and powers of the HC (dotted lines with squares), the maximal L_1- (dashed lines with dots) and L_2- (solid lines with circles) thresholding tests when $p = 20{,}000$ and $n = 100$ with the marginal distribution the standard normal.*

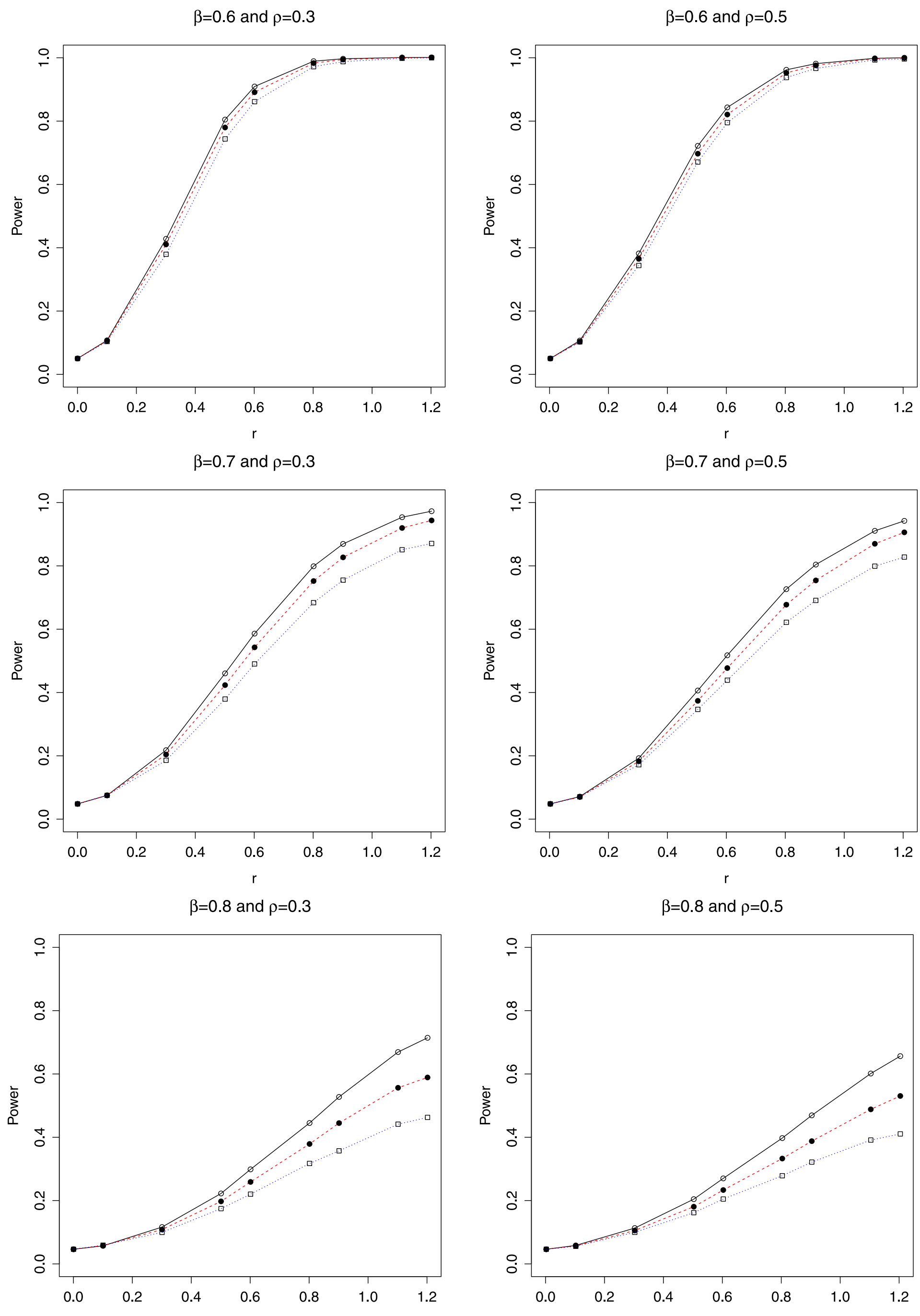

FIG. 3. *Empirical sizes and powers of the HC (dotted lines with squares), the maximal L_1- (dashed lines with dots) and L_2- (solid lines with circles) thresholding tests when $p = 500$ and $n = 100$ with the marginal distribution the standard normal.*

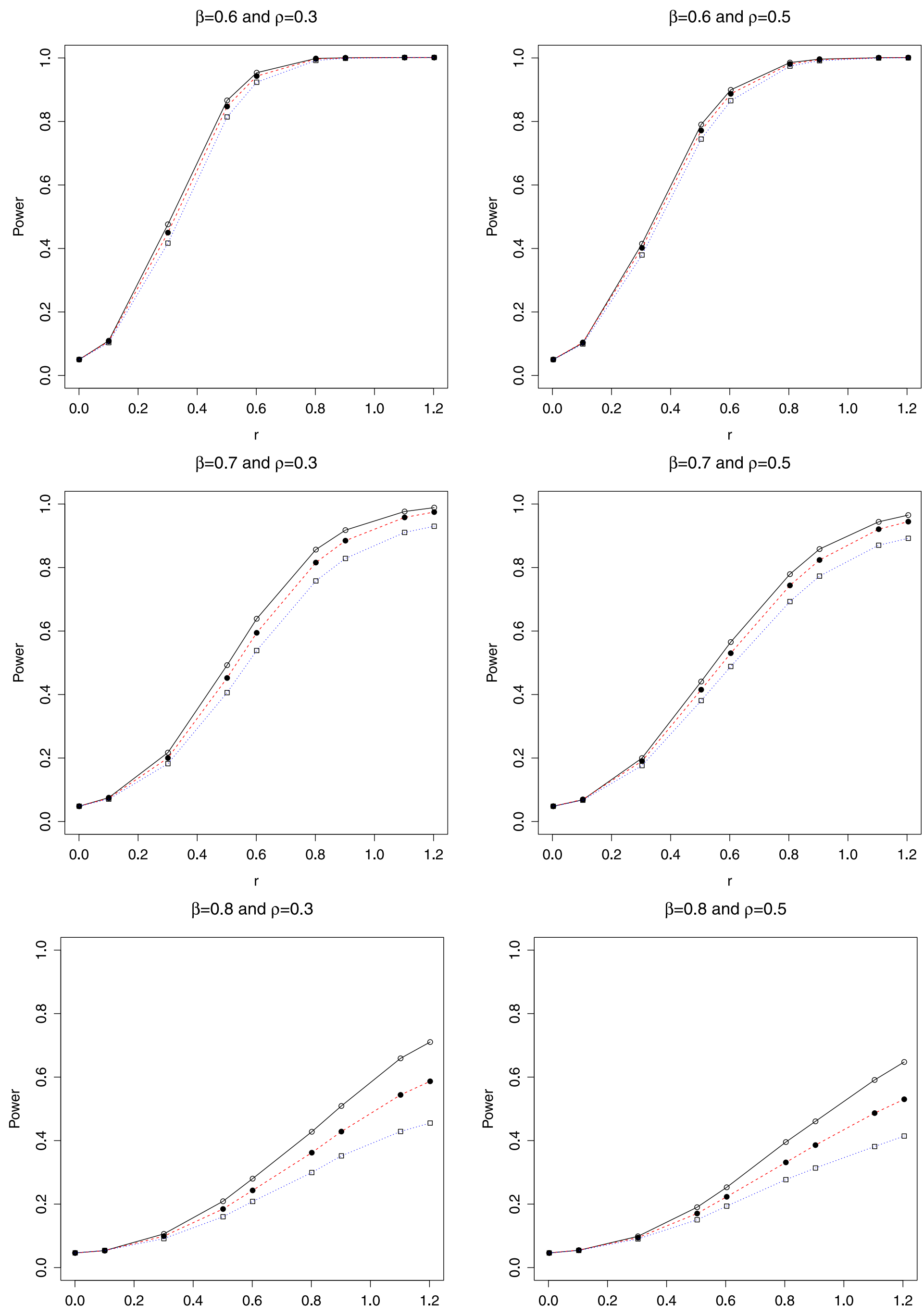

FIG. 4. *Empirical sizes and powers of the HC* (*dotted lines with squares*), *the maximal* L_1- (*dashed lines with dots*) *and* L_2- (*solid lines with circles*) *thresholding tests when* $p = 936$ *and* $n = 200$ *with the marginal distribution the standard normal*.

HC test. We find that for each fixed level of sparsity β, when the signal strength r was increased so that (r, β) moved away from the detection boundary $r = \varrho^*(\beta)$, the difference among the powers of the three tests was enlarged. This was especially the case for the most sparse case of $\beta = 0.8$ and was indeed confirmatory to Theorem 4. The simulated powers of the three tests were very much the same at $r = 0.1$ and were barely changed even when both n and p were increased. This was consistent with the fact that $r = 0.1$ is below the detection boundary for $\beta = 0.7$ and 0.8 considered in the simulation.

6. Discussion. Our analysis shows that there are alternative L_1 and L_2 formulations to the HC test which attain the detection boundary $r = \varrho^*(\beta)$ of the HC test. The tests based on the L_1 and L_2 formulations are more powerful than the HC test when the (r, β) pair is away from the detection boundary such that $r > 2\beta - 1$. The three tests have asymptotically equivalent power when (r, β) is just above the detection boundary.

The detection boundary $r = \varrho^*(\beta)$ coincides with that of the HC test discovered in Donoho and Jin (2004) for the Gaussian data with independent components. That the three tests considered in this paper attain the detection boundary $r = \varrho^*(\beta)$ under the considered sub-Gaussian setting with column-wise dependence can be understood in two aspects. One is that the three test statistics are all directly formulated via the marginal sample means $\bar{X}_j$ which are asymptotically normally distributed; the other is that the thresholding statistics are asymptotically uncorrelated as implied from Proposition 1.

According to Ingster (1997) and Donoho and Jin (2004), $r = \varrho^*(\beta)$ is the optimal detection boundary for Gaussian distributed data with independent components. However, it may not be optimal for the dependent nonparametric setting considered in this paper. Indeed, for weakly dependent Gaussian data, Hall and Jin (2010) showed that the detection boundary $r = \varrho^*(\beta)$ can be lowered by utilizing the dependence. The latter was carried out by pre-transforming the data with $\mathbf{L}$, the inverse of the Cholesky decomposition of $\mathbf{\Sigma}$, or an empirical estimate of $\mathbf{L}$ and then conducting the HC test based on the transformed data. It is expected that the main results of this paper on the relative performance of the three tests would remain valid for the transformed data. Hall and Jin (2008) and Delaigle and Hall (2009) studied the detection boundary for dependent data and Cai and Wu (2012) studied the boundary for detecting mixtures with a general known distribution. However, the optimal detection boundary under the dependent sub-Gaussian distribution setting is still an open problem.

APPENDIX: TECHNICAL DETAILS

In this Appendix we provide proofs to Theorems 2, 3 and 4 reported in Sections 3 and 4. Throughout this Appendix we use $L_p = C \log^b(p)$ to denote slow varying functions for some constant b and positive constant C, and $\phi(\cdot)$ and $\bar{\Phi}(\cdot)$

for the density and survival functions of the standard normal distribution, respectively. Let ρ_k be the correlation coefficient between W_{i1} and $W_{i(k+1)}$, and write $\rho_1 = \rho$ for simplicity and $\mu_j = E(X_{ij})$ for $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, p\}$. Put $\lambda_p(s) = 2s \log p$.

PROOF OF THEOREM 2. Let $u = \bar{\Phi}(\lambda_p^{1/2}(s))$. Write $\mathcal{J}_2(u) := \hat{\mathcal{T}}_{2,n}(s)$ and

$$\mathcal{M}_{2n} = \max_{s \in (0, 1-\eta]} \hat{\mathcal{T}}_{2,n}(s) = \max_{u \in [u_0, 1/2)} \mathcal{J}_2(u),$$

where $u_0 = \bar{\Phi}(\lambda_p^{1/2}(1-\eta))$. Using the same technique for the proof of Theorem 1 in Zhong, Chen and Xu (2013), it may be shown that the joint asymptotic normality of $\mathcal{T}_{2,n}(s)$ at any finite points $\underline{s} = (s_1, \dots, s_d)^T$. This is equivalent to the joint asymptotic normality of $\mathcal{J}_2(u)$ at $u_i = \bar{\Phi}(\sqrt{2s_i \log p})$ for $i = 1, \dots, d$.

We want to show the tightness of the process $\mathcal{J}_2(u)$. Let $f_{n,u}(x) = \sigma_0^{-1}(u) \times x^2 I\{|x| > g(u)\}$, where $g(u) = \bar{\Phi}^{-1}(u)$, $\sigma_0^2(u) = \sigma_0^2(p; s)$ and $\sigma_0^2(p; s) = \sigma^2_{T_{2n},0}(s)/p$. Write

$$\mathcal{J}_2(u) = p^{-1/2} \sum_{j=1}^{p} \{ f_{n,u}(|\sqrt{n}\bar{X}_j|) - E(f_{n,u}(|\sqrt{n}\bar{X}_j|)) \}.$$

Based on the finite dimensional convergence of $\mathcal{J}_2(u)$ and Theorem 1.5.6 in Van der Vaart and Wellner (1996), we only need to show the asymptotically equicontinuous of $\mathcal{J}_2(u)$, that is, for any $\varepsilon > 0$ and $\eta > 0$ there exists a finite partition $\Lambda = \bigcup_{i=1}^k \Lambda_i$ such that

$$\limsup_{n \to \infty} P^* \Big\{ \max_{1 \le i \le k} \sup_{u, v \in \Lambda_i} |\mathcal{J}_2(u) - \mathcal{J}_2(v)| > \varepsilon \Big\} < \eta, \tag{A.1}$$

where P^* is the outer probability measure.

Define $\mathscr{F}_n = \{ f_{n,u}(|\sqrt{n}\bar{X}_j|) = \sigma_0^{-1}(u) |\sqrt{n}\bar{X}_j|^2 I\{|\sqrt{n}\bar{X}_j| > g(u)\} : u \in \Lambda := [u_0, 1/2)\}$ and $\rho(f_{n,u} - f_{n,v}) = [E\{f_{n,u}(|\sqrt{n}\bar{X}_j|) - f_{n,v}(|\sqrt{n}\bar{X}_j|)\}^2]^{1/2}$. It can be shown that if $u > v$,

$$\rho(f_{n,u} - f_{n,v})^2 = \{2 - 2\sigma_0^{-1}(u)\sigma_0(v)\}\{1 + o(1)\}.$$

Thus, for every $\delta_n \to 0$, $\sup_{|u-v|<\delta_n} \rho(f_{n,u} - f_{n,v}) \to 0$, which implies that for each $\delta > 0$, Λ can be partitioned into finitely many sets $\Lambda_1, \dots, \Lambda_k$ satisfying

$$\max_{1 \le i \le k} \sup_{u, v \in \Lambda_i} \rho(f_{n,u} - f_{n,v}) < \delta.$$

Let $N_0 := N(\varepsilon, \mathscr{F}_n, \rho)$ be the bracketing number, the smallest number of functions $f_1, \dots, f_{N_0}$ in $\mathscr{F}_n$ such that for each f in $\mathscr{F}_n$ there exists an f_i ($i \in \{1, \dots, N_0\}$) satisfying $\rho(f - f_i) \le \varepsilon \le 1$. Applying Theorem 2.2 in Andrews and Pollard

(1994), if the following two conditions hold for an even integer $Q \geq 2$ and a real number $\gamma > 0$ such that

$$\sum_{d=1}^{\infty} d^{Q-2}\alpha(d)^{\gamma/(Q+\gamma)} < \infty \quad \text{and} \tag{A.2}$$

$$\int_0^1 \varepsilon^{-\gamma/(2+\gamma)} N(\varepsilon, \mathscr{F}_n, \rho)^{1/Q}\, d\varepsilon < \infty, \tag{A.3}$$

we have for n large enough $\| \sup_{\substack{\rho(f_{n,u}-f_{n,v})<\delta \\ u,v\in\Lambda_i}} |\mathcal{J}_2(u) - \mathcal{J}_2(v)| \|_Q < k^{-1/Q}\eta\varepsilon$.

Invoking the maximal inequality of Pisier (1983), it follows that

$$\Big\| \max_{1\leq i\leq k} \sup_{\substack{\rho(f_{n,u}-f_{n,v})<\delta \\ s,t\in\Lambda_i}} |\mathcal{J}_2(u) - \mathcal{J}_2(v)| \Big\|_Q < \eta\varepsilon.$$

Now using the Markov inequality, we get for n large enough

$$\begin{aligned} &P^*\Big\{ \max_{1\leq i\leq k} \sup_{u,v\in\Lambda_i} |\mathcal{J}_2(u) - \mathcal{J}_2(v)| > \varepsilon \Big\} \\ &\quad \leq \Big\| \max_{1\leq i\leq k} \sup_{\substack{\rho(f_{n,u}-f_{n,v})<\delta \\ u,v\in\Lambda_i}} |\mathcal{J}_2(u) - \mathcal{J}_2(v)| \Big\|_Q / \varepsilon < \eta. \end{aligned}$$

Hence, the condition (A.1) holds and $\mathcal{J}_2(u)$ is asymptotically tight.

It remains to show (A.2) and (A.3) hold. For (A.3), we note that $\mathscr{F}_n$ is a V-C class for each n. This is because

$$\mathscr{G}_n = \{ f_{n,u}(x) = \sigma_0^{-1}(u) I(x > g(u)) : u \in (u_0, 1/2) \}$$

is a V-C class with VC index 2. Let $\varphi(x) = x^2$. Then $\mathscr{F}_n = \varphi \cdot \mathscr{G}_n$ is a V-C class by Lemma 2.6.18 in Van der Vaart and Wellner (1996). Let $G_n(x, u_0) = \sup_{u\in\Lambda} |f_{n,u}(x)|$ be the envelop function for class $\mathscr{F}_n$. Clearly, we can take $G_n(x, u_0) = \sigma_0^{-1}(u_0)x^2$. It is easy to see that $\rho\{G_n(|\sqrt{n}\bar{X}_i|, u_0)\} < \infty$ for a constant $u_0 > 0$. Applying a result on covering number of V-C classes [Theorem 2.6.7, Van der Vaart and Wellner (1996)], we get $N(\varepsilon, \mathscr{F}_n, \rho) \leq K\varepsilon^{-2}$ for a universal constant K. It can be verified that if $Q > 2 + \gamma$, then (A.3) holds. The condition (A.2) follows from the assumption that $\rho_Z(d) \leq C\alpha^d$.

As a result, $\mathcal{J}_2(u)$ converge to a zero mean Gaussian process $\mathcal{N}_2(u)$ with

$$\operatorname{Cov}(\mathcal{N}_2(u), \mathcal{N}_2(v)) = \frac{\sigma_0(u)}{\sigma_0(v)} = \exp\Big(-\frac{1}{2}[\log\{\sigma_0^2(v)\} - \log\{\sigma_0^2(u)\}]\Big)$$

for $u < v$. It can be shown that there exists an Ornstein–Uhlenbeck (O–U) process $\mathcal{U}_2(\cdot)$ with mean zero 0 and $E(\mathcal{U}_2(u)\mathcal{U}_2(v)) = \exp\{-|u - v|\}$ such that

$\mathcal{N}_2(u) = \mathcal{U}_2(\frac{1}{2}\log\{\sigma_0^2(u)\})$. Therefore, by a result for the O–U process in Leadbetter, Lindgren and Rootzén [(1983), page 217],

$$P\Big(\max_{s\in\mathcal{S}} \hat{\mathcal{T}}_{2,n}(s) < B_{\tau_n}(x)\Big) = P\Big(\max_{u\in\Lambda} \hat{\mathcal{N}}_2(u) < B_{\tau_n}(x)\Big)\{1+o(1)\}$$
$$= P\Big(\max_{u\in(0,\tau_n)} \mathcal{U}_2(u) < B_{\tau_n}(x)\Big) \to \exp\{-\exp(-x)\},$$

where $\tau_n = \frac{1}{2}\log\{\sigma_0^2(\frac{1}{2})/\sigma_0^2(u_0)\}$, $B_{\tau_n}(x) = (x + b^*(\tau_n))/a(\tau_n)$, $a(t) = (2\log(t))^{1/2}$ and $b^*(t) = 2\log(t) + 2^{-1}\log\log(t) - \frac{1}{2}\log(\pi)$. From (2.4), we have $\tau_n = \frac{1-\eta}{2}\log p\{1+o(1)\}$. Since

$$a(\tau_n)\max_{u\in(0,\tau_n)} \mathcal{U}_2(u) - b^*(\tau_n) = \frac{a(\tau_n)}{a(\log p)}\Big[a(\log p)\max_{u\in(0,\tau_n)} \mathcal{U}_2(u) - b^*(\log p)\Big]$$
$$+ \frac{a(\tau_n)}{a(\log p)} b^*(\log p) - b^*(\tau_n),$$

$a(\tau_n)/a(\log p) \to 1$ and

$$\frac{a(\tau_n)}{a(\log p)} b^*(\log p) - b^*(\tau_n) = \frac{a(\tau_n)}{a(\log p)}[b^*(\log p) - b^*(\tau_n)]$$
$$+ b^*(\tau_n)\Big[\frac{a(\tau_n)}{a(\log p)} - 1\Big] \to -\log\frac{(1-\eta)}{2},$$

we have

$$a(\tau_n)\max_{u\in(0,\tau_n)} \mathcal{U}_2(u) - b^*(\tau_n)$$
$$= a(\log p)\max_{u\in(0,\tau_n)} \mathcal{U}_2(u) - \Big(b^*(\log p) + \log\frac{(1-\eta)}{2}\Big).$$

Finally, note that $b^*(\log p) + \log\frac{(1-\eta)}{2} = b(\log p, \eta)$. This finishes the proof of Theorem 2. □

PROOF OF THEOREM 3. (i). The proof is made under four cases. For each case, we find the corresponding detectable region and the union of the four regions are the overall detectable region of the thresholding test. Basically, we show for any (β, r) above $\varrho^*(\beta)$ within one of the four cases, there exists at least one threshold level s such that H_1 is detectable. For notation simplification, we only keep the leading order terms for $\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s)$, $\sigma_{T_{2n,1}}(s)$, $\sigma_{T_{2n,0}}(s)$ and $\Delta_2(s; r, \beta)$.

Case 1: $s \le r$ and $s \le \beta$. In this case, $\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s) = L_p p^{1-\beta}$ and $\sigma_{T_{2n},1}(s) = \sigma_{T_{2n},0}(s) = L_p p^{(1-s)/2}$. Hence,

$$\Delta_2(s; r, \beta) = \frac{\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s)}{\sigma_{T_{2n},1}(s)} = L_p p^{(1+s-2\beta)/2}.$$

So to make $(\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s))/\sigma_{T_{2n},1}(s) \to \infty$, we need $s > 2\beta - 1$. It follows that the detectable region for this case is $r \geq 2\beta - 1$. Specifically, if we select $s = \min\{r, \beta\}$, we arrive at the best divergence rate for $\Delta_2(s; r, \beta)$ of order $L_p p^{(1+\min\{r,\beta\}-2\beta)/2}$.

Case 2: $s \leq r$ and $s > \beta$. In this case, $\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s) = L_p p^{1-\beta}$, $\sigma_{T_{2n},1}(s) = L_p p^{(1-\beta)/2}$, and $\sigma_{T_{2n},0}(s) = L_p p^{(1-s)/2}$. Then,

$$\Delta_2(s; r, \beta) = \frac{\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s)}{\sigma_{T_{2n},1}(s)} = L_p p^{(1-\beta)/2}.$$

So the detectable region in the (β, r) plane is $r > \beta$. In this region, the best divergence rate of Δ_2 is of order $L_p p^{(1-\beta)/2}$ for any $\beta < s \leq r$.

Case 3: $s > r$ and $s \leq (\sqrt{s} - \sqrt{r})^2 + \beta$. The case is equivalent to $\sqrt{r} < \sqrt{s} \leq (r + \beta)/(2\sqrt{r})$ and $\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s) = L_p p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}$, $\sigma_{T_{2n},1}(s) = \sigma_{T_{2n},0} = L_p p^{(1-s)/2}$. Then

$$\Delta_2(s; r, \beta) = \frac{\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s)}{\sigma_{T_{2n},1}(s)} = L_p p^{1/2-\beta+r-(\sqrt{s}-2\sqrt{r})^2/2}. \tag{A.4}$$

To ensure (A.4) diverging to infinity, we need

$$2\sqrt{r} - \sqrt{1 - 2\beta + 2r} < \sqrt{s} < 2\sqrt{r} + \sqrt{1 - 2\beta + 2r}.$$

Thus, the detectable region must satisfy

$$\sqrt{r} < (r + \beta)/(2\sqrt{r}), \qquad 1 - 2\beta + 2r > 0 \quad \text{and}$$
$$2\sqrt{r} - \sqrt{1 - 2\beta + 2r} \leq (r + \beta)/(2\sqrt{r}).$$

This translates to

$$\beta - \tfrac{1}{2} < r < \beta \quad \text{and either } r \leq \beta/3 \text{ or } r > \beta/3 \text{ and } \quad r \geq (1 - \sqrt{1 - \beta})^2.$$

Case 4: $s > r$ and $s > (\sqrt{s} - \sqrt{r})^2 + \beta$. This is equivalent to $\sqrt{s} > \max\{(r + \beta)/(2\sqrt{r}), \sqrt{r}\}$. In this case, $\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s) = L_p p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}$, $\sigma_{T_{2n},1}(s) = L_p p^{(1-(\sqrt{s}-\sqrt{r})^2-\beta)/2}$. Then

$$\Delta_2(s; r, \beta) = \frac{\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s)}{\sigma_{T_{2n},1}(s)} = L_p p^{(1-(\sqrt{s}-\sqrt{r})^2-\beta)/2}.$$

Hence, it requires that

$$\sqrt{r} - \sqrt{1 - \beta} < \sqrt{s} < \sqrt{r} + \sqrt{1 - \beta}.$$

In order to find an s, we need $\sqrt{r} + \sqrt{1 - \beta} > \max\{(r + \beta)/(2\sqrt{r}), \sqrt{r}\}$. If $\sqrt{r} > (r + \beta)/(2\sqrt{r})$, namely, $r > \beta$, the above inequality is obviously true. If $r \leq \beta$, then $\sqrt{r} + \sqrt{1 - \beta} > (r + \beta)/(2\sqrt{r})$ is equivalent to $r > (1 - \sqrt{1 - \beta})^2$. So the detectable region is $r > (1 - \sqrt{1 - \beta})^2$ in this case.

In summary of cases 1–4, the union of the detectable regions in the above four cases is $r > \varrho^*(\beta)$, as illustrated in Figure 5.

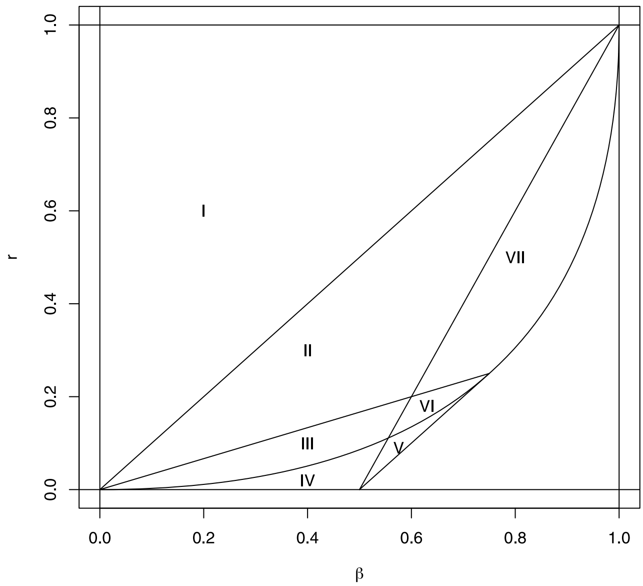

FIG. 5. *The detectable subregions of the* L_2 *threshold test. Case* 1: *the union of* {*I*, *II*, *III*, *IV*}; *Case* 2, *the region is I*; *Case* 3: *the union of* {*II*, *III*, *IV*, *V*, *VI*, *VII*}; *Case* 4: *the union of* {*I*, *II*, *III*, *VI*, *VII*}.

Now we are ready to prove the theorem. We only need to show that the sum of type I and II errors of the maximal test goes to 0 when $r > \varrho^*(\beta)$. Because the maximal test is of asymptotic α_n level, it suffices to show that the power goes to 1 in the detectable region as $n \to \infty$ and $\alpha_n \to 0$. Recall that the α_n level rejection region is $R_{\alpha_n} = \{\hat{\mathcal{M}}_{2n} > \mathcal{B}_{\alpha_n}\}$. From Theorem 2, we notice that $\mathcal{B}_{\alpha_n} = O\{(\log\log p)^{1/2}\} := L_p^*$. Then, it is sufficient if

$$P(\mathcal{M}_{2n}/L_p^* \to \infty) \to 1 \qquad \text{as } n \to \infty \tag{A.5}$$

at every (β, r) in the detectable region. Since $\mathcal{M}_{2n} \geq \mathcal{T}_{2n}(s)$ for any $s \in \mathcal{S}$, therefore, (A.5) is true if for any point in the detectable region, there exists a $\lambda_p(s) = 2s \log p$ such that

$$\mathcal{T}_{2n}(s)/L_p^* \xrightarrow{p} \infty. \tag{A.6}$$

Therefore, we want to show

$$\begin{aligned} &\frac{T_{2n}(s) - \mu_{T_{2n},0}(s)}{L_p^* \sigma_{Tn,0}(s)} \\ &\quad = \left(\frac{T_{2n}(s) - \mu_{T_{2n},1}(s)}{L_p^* \sigma_{T_{2n},1}(s)} + \frac{\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s)}{L_p^* \sigma_{T_{2n},1}(s)} \right) \frac{\sigma_{T_{2n},1}(s)}{\sigma_{T_{2n},0}(s)} \\ &\quad \xrightarrow{p} \infty. \end{aligned} \tag{A.7}$$

Because $(T_{2n}(s) - \mu_{T_{2n},1}(s))/L_p^*\sigma_{T_{2n},1}(s) = o_p(1)$ and $\sigma_{T_{2n},0}(s) \le \sigma_{T_{2n},1}(s)$, (A.7) is true if $(\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s))/L_p^*\sigma_{T_{2n},1}(s) \to \infty$. As we have shown in the early proof, for every (r, β) in the detectable region, there exists an s such that $\frac{\mu_{T_{2n},1}(s)-\mu_{T_{2n},0}(s)}{L_p\sigma_{T_{2n},1}(s)} \to \infty$ for any slow varying function L_p. This concludes (A.6) and hence (A.5), which completes the proof of part (i).

(ii) Note that

$$\hat{\mathcal{M}}_{2n} = \max_{s\in\mathcal{S}_n}\biggl\{\bigl(\mathcal{T}_{2n,1}(s)R_2(s) + \Delta_{2,0}(s; r, \beta)\bigr)\tilde{e}_2(s) + \frac{\mu_{T_{\gamma n},0}(s) - \hat{\mu}_{T_{\gamma n},0}(s)}{\tilde{\sigma}_{T_{\gamma n},0}(s)}\biggr\},$$

where $R_2(s)$, $\tilde{e}_2(s)$ and $\mathcal{T}_{2n,1}(s)$ are defined in (4.4) and

$$\begin{aligned}\Delta_{2,0}(s; r, \beta) &= \frac{\mu_{T_{2n},1}(s) - \mu_{T_{2n},0}(s)}{\sigma_{T_{2n},0}(s)}\\ &= (s\pi \log p)^{1/4}(r/s)p^{1/2-\beta+s/2}I(r > s)\\ &\quad + \frac{s^{1/4}(\pi \log p)^{-1/4}}{2(\sqrt{s}-\sqrt{r})}p^{1/2-(\sqrt{s}-\sqrt{r})^2-\beta+s/2}I(r < s).\end{aligned} \tag{A.8}$$

If $r < \varrho^*(\beta)$, then $r < \beta$ and $r < (r + \beta)^2/(4r)$. Hence,

$$R_2(s) = \begin{cases} 1 + o(1), & \text{if } s \le r;\\ 1 + o(1), & \text{if } r < s \le \dfrac{(r+\beta)^2}{4r};\\ s^{1/4}(\sqrt{s}-\sqrt{r})^{-1/2}p^{1/2(2\sqrt{sr}-r-\beta)}\{1 + o(1)\}, & \text{if } s > \dfrac{(r+\beta)^2}{4r}.\end{cases}$$

It is also noticed that $r < \varrho^*(\beta)$ implies that $(r + \beta)^2/(4r) > 1$. Therefore, for all $s \in \mathcal{S}_n$, $R_2(s) = 1 + o(1)$.

If $r < \varrho^*(\beta)$, then $r < 2\beta - 1$. Hence, $\max_{s\le r} \Delta_{2,0}(s; r, \beta) \le L_p p^{1/2-\beta+r/2} \to 0$ as $p(n) \to \infty$.

If $r < \varrho^*(\beta)$ and $r < 1/4$, then $r < \beta - 1/2$. It follows that, for all $s > r$,

$$1/2 - (\sqrt{s}-\sqrt{r})^2 - \beta + s/2 = 1/2 + r - \beta - \tfrac{1}{2}(\sqrt{s} - 2\sqrt{r})^2 \le 1/2 + r - \beta < 0.$$

If $r < \varrho^*(\beta)$ and $r > 1/4$, then for all $s > r$,

$$1/2 - (\sqrt{s}-\sqrt{r})^2 - \beta + s/2 \le 1/2 + r - \beta - \tfrac{1}{2}(1 - 2\sqrt{r})^2 < 0.$$

Hence, $\max_{s>r} \Delta_{2,0}(s; r, \beta) \le L_p p^{1/2+r-\beta}I\{r < 1/4\} + L_p p^{1-\beta-(1-\sqrt{r})^2}I\{r > 1/4\} \to 0$ as $p(n) \to \infty$. In summary, we have $R_2(s) = 1 + o(1)$ and $\max_{s\in\mathcal{S}_n} \Delta_{2,0}(s; r, \beta) \to 0$ if $r < \varrho^*(\beta)$. Therefore, together with assumption (2.6), $\hat{\mathcal{M}}_{2n} = \max_{s\in\mathcal{S}_n} \mathcal{T}_{2n,1}(s)\{1 + o_p(1)\}$.

We note that, by employing the same argument of Theorem 2, it can be shown that

$$P\Bigl(a(\log p)\max_{s\in\mathcal{S}} \mathcal{T}_{2n,1}(s) - b(\log p, \delta) \le x\Bigr) \to \exp(-e^{-x}),$$

where δ is defined just above (A.11). Then the power of the test

$$\begin{aligned}
&P(\hat{\mathcal{M}}_{2n} > (\mathcal{E}_{\alpha_n} + b(\log p, \eta))/a(\log p)) \\
&\quad = P(\hat{\mathcal{M}}_{2n} > (\mathcal{E}_{\alpha_n} + b(\log p, \delta))/a(\log p))\{1 + o(1)\} \\
&\quad = \alpha_n\{1 + o(1)\} \to 0.
\end{aligned}$$

Thus, the sum of type I and II errors goes to 1. This completes the proof of part (ii). $\square$

PROOF OF THEOREM 4. We first prove that $\hat{\mathcal{M}}_{\gamma n} \sim \max_{s\in\mathcal{S}_n} \Delta_{\gamma,0}(s; r, \beta)$, which will be proved in two parts:

$$\hat{\mathcal{M}}_{\gamma n} \sim \mathcal{M}_{\gamma n} \quad \text{and} \tag{A.9}$$

$$\mathcal{M}_{\gamma n} \sim \max_{s\in\mathcal{S}_n} \Delta_{\gamma,0}(s; r, \beta), \tag{A.10}$$

where $\mathcal{M}_{\gamma n} = \max_{s\in\mathcal{S}_n} \mathcal{T}_{\gamma n}(s) = \max_{s\in\mathcal{S}_n}\{\mathcal{T}_{\gamma n,1}(s)R_\gamma(s) + \Delta_{\gamma,0}(s; r, \beta)\}$.

To show (A.9), note the decomposition for $\hat{\mathcal{M}}_{\gamma n}$ in (4.4). Let $\tilde{\mathcal{M}}_{\gamma n} = \max_{s\in\mathcal{S}_n}\{\mathcal{T}_{\gamma n}(s)\tilde{e}_\gamma(s)\}$. We can first show that $\hat{\mathcal{M}}_{\gamma n} \sim \tilde{\mathcal{M}}_{\gamma n}$ because of the following inequality:

$$\begin{aligned}
&\tilde{\mathcal{M}}_{\gamma n} - \left|\max_{s\in\mathcal{S}_n} \frac{\mu_{T_{\gamma n},0}(s) - \hat{\mu}_{T_{\gamma n},0}(s)}{\tilde{\sigma}_{T_{\gamma n},0}(s)}\right| \\
&\qquad \le \hat{\mathcal{M}}_{\gamma n} \le \tilde{\mathcal{M}}_{\gamma n} + \left|\max_{s\in\mathcal{S}_n} \frac{\mu_{T_{\gamma n},0}(s) - \hat{\mu}_{T_{\gamma n},0}(s)}{\tilde{\sigma}_{T_{\gamma n},0}(s)}\right|.
\end{aligned}$$

Under condition (2.6), that is, $\max_{s\in\mathcal{S}} \tilde{\sigma}^{-1}_{T_{\gamma n},0}(s)(\mu_{T_{\gamma n},0}(s) - \hat{\mu}_{T_{\gamma n},0}(s)) = o(1)$, hence, $\hat{\mathcal{M}}_{\gamma n} \sim \tilde{\mathcal{M}}_{\gamma n}$. Second, we can show $\mathcal{M}_{\gamma n} \sim \tilde{\mathcal{M}}_{\gamma n}$. Note the following inequality:

$$\begin{aligned}
&\min\Big\{\mathcal{M}_{\gamma n} \min_{s\in\mathcal{S}_n} \tilde{e}_\gamma(s), \mathcal{M}_{\gamma n} \max_{s\in\mathcal{S}_n} \tilde{e}_\gamma(s)\Big\} \\
&\qquad \le \tilde{\mathcal{M}}_{\gamma n} \le \max\Big\{\mathcal{M}_{\gamma n} \min_{s\in\mathcal{S}_n} \tilde{e}_\gamma(s), \mathcal{M}_{\gamma n} \max_{s\in\mathcal{S}_n} \tilde{e}_\gamma(s)\Big\}.
\end{aligned}$$

Under conditions (C.1)–(C.4), $\min_{s\in\mathcal{S}_n} \tilde{e}_\gamma(s) = \max_{s\in\mathcal{S}_n} \tilde{e}_\gamma(s) = 1 + o(1)$. So we have

$$\tilde{\mathcal{M}}_{\gamma n} \sim \mathcal{M}_{\gamma n} \min_{s\in\mathcal{S}_n} \tilde{e}_\gamma(s) \sim \mathcal{M}_{\gamma n} \min_{s\in\mathcal{S}_n} \tilde{e}_\gamma(s) \sim \mathcal{M}_{\gamma n}.$$

In summary, we have $\hat{\mathcal{M}}_{\gamma n} \sim \tilde{\mathcal{M}}_{\gamma n} \sim \mathcal{M}_{\gamma n}$. Therefore, $\hat{\mathcal{M}}_{\gamma n} \sim \mathcal{M}_{\gamma n}$.

The path leading to (A.10) is the following. First of all, it can be shown using an argument similar to the one used in the proof of Theorem 2 that

$$P\Big(a(\log p) \max_{s\in\mathcal{S}} \mathcal{T}_{\gamma n,1}(s) - b(\log p, \delta) \le x\Big) \to \exp(-e^{-x}),$$

where $\delta = \max\{\eta - r + 2r\sqrt{1-\eta} - \beta, \eta\} I(r < 1-\eta) + \max\{1-\beta, \eta\} I(r > 1-\eta)$. Thus, for $\gamma = 0, 1$ and 2,

$$\max_{s \in \mathcal{S}} \mathcal{T}_{\gamma n,1}(s) = O_p\{\log^{1/2}(\log p)\}. \tag{A.11}$$

Equations (A.13) to (A.20) in the following reveal that for all $s \in \mathcal{S}$ and $r > \varrho^*(\beta)$, we can classify $s \in \mathcal{S}$ into two sets $\mathcal{S}_1$ and $\mathcal{S}_2$ such that

(i) $\Delta_{\gamma,0}(s; r, \beta) \gg R_\gamma(s)$ for $s \in \mathcal{S}_1$

(ii) $\Delta_{\gamma,0}(s; r, \beta) \to 0$ and $R_\gamma(s) = 1 + o(1)$ for $s \in \mathcal{S}_2$,

where "$c \gg d$" means that $c/d = L_p p^\xi$ for some $\xi > 0$. Because r is above the detection boundary $\varrho^*(\beta)$, there exists at least one $s \in \mathcal{S}_1$ such that $\Delta_{\gamma,0}(s; r, \beta) \to \infty$. Hence,

$$\max_{s \in \mathcal{S}} \Delta_{\gamma,0}(s; r, \beta) = \max_{s \in \mathcal{S}_1} \Delta_{\gamma,0}(s; r, \beta) \gg \max_{s \in \mathcal{S}} R_\gamma(s). \tag{A.12}$$

Namely, the maximum of $\Delta_{\gamma,0}(s; r, \beta)$ is reached on Set $\mathcal{S}_1$ where $\Delta_{\gamma,0}(s; r, \beta)$ diverges at a much faster rate than that of $\tilde{R}_\gamma(s)$, if the latter ever diverges.

Let $A(s) = \mathcal{T}_{2n,1}(s) R_\gamma(s)$. Combining (A.11) and (A.12), we have

$$\Bigl|\max_{s \in \mathcal{S}_n} \mathcal{T}_{\gamma n,1}(s)\Bigr| \Bigl|\max_{s \in \mathcal{S}_n} R_\gamma(s)\Bigr| = o_p\Bigl\{\max_{s \in \mathcal{S}_n} \Delta_{\gamma,0}(s; r, \beta)\Bigr\}.$$

This implies that $|\max_{s \in \mathcal{S}_n} A(s)| = o_p\{\max_{s \in \mathcal{S}_n} \Delta_{\gamma,0}(s; r, \beta)\}$. Together with the following inequality:

$$\begin{aligned}\max_{s \in \mathcal{S}_n} \Delta_{\gamma,0}(s; r, \beta) - \Bigl|\max_{s \in \mathcal{S}_n} A(s)\Bigr| &\le \max_{s \in \mathcal{S}_n}\bigl\{A(s) + \Delta_{\gamma,0}(s; r, \beta)\bigr\} \\ &\le \max_{s \in \mathcal{S}_n} \Delta_{\gamma,0}(s; r, \beta) + \max_{s \in \mathcal{S}_n} A(s);\end{aligned}$$

we conclude that (A.10) holds.

It remains to show the existence of $\mathcal{S}_1$ and $\mathcal{S}_2$ in arriving at (A.12). We only prove it for the L_2 test. To complete that, we compare the relative order between $\Delta_{2,0}(s; r, \beta)$ and $R_2(s)$ for three regions above the detection boundary $\varrho^*(\beta)$: (i) $r > \beta$ (ii) $r \in (2\beta - 1, \beta]$ and (iii) $r \in (\varrho^*(\beta), 2\beta - 1]$. In regions (i) and (ii) with $r > (1 - \sqrt{1-\beta})^2$, we can show that

$$\Delta_{2,0}(s; r, \beta) \gg R_2(s) \quad \text{for } s > 2\beta - 1; \tag{A.13}$$

$$\Delta_{2,0}(s; r, \beta) \to 0 \quad \text{and} \quad R_2(s) = 1 + o(1) \quad \text{for } s \le 2\beta - 1. \tag{A.14}$$

In region (ii) with $r < (1 - \sqrt{1-\beta})^2$, we have

$$\Delta_{2,0}(s; r, \beta) \gg R_2(s) \quad \text{for } 2\beta - 1 < s \le (2\sqrt{r} + \sqrt{1 + 2r - 2\beta})^2, \tag{A.15}$$

$$\begin{aligned}&\Delta_{2,0}(s; r, \beta) \to 0 \quad \text{and} \quad R_2(s) = 1 + o(1) \quad \text{for } s \le 2\beta - 1 \\ &\qquad\qquad \text{and } (2\sqrt{r} + \sqrt{1 + 2r - 2\beta})^2 < s < 1.\end{aligned} \tag{A.16}$$

For $r \in (\varrho^*(\beta), 2\beta - 1]$ in region (iii). If $r > (1 - \sqrt{1-\beta})^2$, define $D_1 = (0, (2\sqrt{r} - \sqrt{1+2r-2\beta})^2)$ and $D_2 = ((2\sqrt{r} - \sqrt{1+2r-2\beta})^2, 1)$. Then it may be shown that

$$\Delta_{2,0}(s; r, \beta) \to 0 \quad \text{and} \quad R_2(s) = 1 + o(1) \qquad \text{for } s \in D_1; \tag{A.17}$$

$$\Delta_{2,0}(s; r, \beta) \gg R_2(s) \qquad \text{for } s \in D_2. \tag{A.18}$$

If $r < (1 - \sqrt{1-\beta})^2$, define $D_3 = (0, (2\sqrt{r} - \sqrt{1+2r-2\beta})^2) \cup ((2\sqrt{r} + \sqrt{1+2r-2\beta})^2, 1)$ and $D_4 = ((2\sqrt{r} - \sqrt{1+2r-2\beta})^2, (2\sqrt{r} + \sqrt{1+2r-2\beta})^2)$. Then, it can be shown that

$$\Delta_{2,0}(s; r, \beta) \to 0 \quad \text{and} \quad R_2(s) = 1 + o(1) \qquad \text{for } s \in D_3; \tag{A.19}$$

$$\Delta_{2,0}(s; r, \beta) \gg R_2(s) \qquad \text{for } s \in D_4. \tag{A.20}$$

The results in (A.13)–(A.20) indicate that in each region listed above, $\max \Delta_{2,0}(s; r, \beta)$ will be attained in situations covered by (A.13), (A.15), (A.18) and (A.20), which together imply (A.12).

Next, we compute $\Delta_{\gamma,0}(s; r, \beta)$ for the HC ($\gamma = 0$) and the L_1 ($\gamma = 1$) test. For the HC test, let $G_{p,1}(s) = P(Y_{i,n} > 2s \log p)$. Under assumptions (C.1)–(C.2), applying the large deviation results [Petrov (1995)], it may be shown that

$$G_{p,1}(s) = \{(2\sqrt{\pi \log p}(\sqrt{s} - \sqrt{r}))^{-1} p^{-(\sqrt{s}-\sqrt{r})^2}\}\{1 + o(1)\} \qquad \text{if } r < s \text{ and}$$

$$G_{p,1}(s) = \{1 - (2\sqrt{\pi \log p}(\sqrt{r} - \sqrt{s}))^{-1} p^{-(\sqrt{r}-\sqrt{s})^2}\}\{1 + o(1)\} \qquad \text{if } r > s.$$

The mean and variance of $T_{0n}(s)$ under H_0 are $\mu_{T_{0n},0}(s) = (\sqrt{s\pi \log p})^{-1} \times p^{1-s}\{1 + o(1)\}$ and $\sigma^2_{T_{0n},0}(s) = (\sqrt{s\pi \log p})^{-1} p^{1-s}\{1 + o(1)\}$ respectively. The mean and variance of $T_{0n}(s)$ under the H_1 as specified in (C.4) are, respectively,

$$\mu_{T_{0n},1}(s) = p^{1-\beta} G_{p,1}(s) + (p - p^{1-\beta}) 2\bar{\Phi}(\lambda_p^{1/2}(s))\{1 + o(1)\} \qquad \text{and}$$

$$\begin{aligned} \sigma^2_{T_{0n},1}(s) &= p^{1-\beta} G_{p,1}(s)(1 - G_{p,1}(s)) \\ &\quad + p(1 - p^{-\beta}) 2\bar{\Phi}(\lambda_p^{1/2}(s))(1 - 2\bar{\Phi}(\lambda_p^{1/2}(s))). \end{aligned}$$

These imply that, up to a factor $\{1 + o(1)\}$,

$$\begin{aligned} &\mu_{T_{0n},1}(s) - \mu_{T_{0n},0}(s) \\ &\quad = \{(2\sqrt{\pi \log p}(\sqrt{s} - \sqrt{r}))^{-1} p^{1-\beta-(\sqrt{s}-\sqrt{r})^2} I(r < s) \\ &\qquad\qquad + p^{1-\beta} I(r > s)\} \end{aligned} \tag{A.21}$$

and

$$R_0(s) = \begin{cases} 1, & \text{if } s \le (\sqrt{s} - \sqrt{r})^2 + \beta; \\ s^{1/4} |2(\sqrt{s} - \sqrt{r})|^{-1/2} p^{-1/2((\sqrt{s}-\sqrt{r})^2 + \beta - s)}, & \\ \qquad \text{if } s > (\sqrt{s} - \sqrt{r})^2 + \beta. & \end{cases}$$

Hence,

$$\begin{aligned}\Delta_{0,0}(s;r,\beta) &= \frac{s^{1/4}}{2(\sqrt{s}-\sqrt{r})(\pi\log p)^{1/4}}p^{1/2-\beta-(\sqrt{s}-\sqrt{r})^2+s/2}I(r<s)\\ &\quad+(s\pi\log p)^{1/4}p^{1/2-\beta+s/2}I(r>s).\end{aligned}\tag{A.22}$$

For the L_1 test, the mean and variances of $T_{1n}(s)$ under H_1 specified in (C.4) are, respectively, up to a factor $1+o(1)$,

$$\begin{aligned}\mu_{T_{1n},1}(s) &= \frac{\sqrt{s}}{\sqrt{2\pi}(\sqrt{s}-\sqrt{r})}p^{1-\beta-(\sqrt{s}-\sqrt{r})^2}I(r<s)\\ &\quad+(\sqrt{2r\log p})p^{1-\beta}I(r>s)+\sqrt{2/\pi}\,p^{1-s} \qquad \text{and}\\ \sigma^2_{T_{1n},1}(s) &= \frac{s\sqrt{\log p}}{\sqrt{\pi}(\sqrt{s}-\sqrt{r})}p^{1-\beta-(\sqrt{s}-\sqrt{r})^2}I(r<s)+p^{1-\beta}I(r>s)\\ &\quad+2\sqrt{(s/\pi)\log p}\,p^{1-s}.\end{aligned}$$

It follows that, up to a factor $1+o(1)$,

$$\begin{aligned}\mu_{T_{1n},1}(s)-\mu_{T_{1n},0}(s) &= \frac{\sqrt{s}}{\sqrt{2\pi}(\sqrt{s}-\sqrt{r})}p^{1-\beta-(\sqrt{s}-\sqrt{r})^2}I(r<s)\\ &\quad+(\sqrt{2r\log p})p^{1-\beta}I(r>s)\end{aligned}\tag{A.23}$$

and

$$R_1(s)=\begin{cases}1, & \text{if } s\le r \text{ and } s\le\beta;\\ (\sqrt{2})^{-1}\left(\dfrac{s}{\pi}\right)^{-1/4}(\log p)^{-1/4}p^{(s-\beta)/2}, & \\ & \text{if } s\le r \text{ and } s\ge\beta;\\ 1, & \text{if } s>r \text{ and } s\le(\sqrt{s}-\sqrt{r})^2+\beta;\\ s^{1/4}(2\sqrt{s}-2\sqrt{r})^{-1/2}p^{-1/2((\sqrt{s}-\sqrt{r})^2+\beta-s)}, & \\ & \text{if } s>r \text{ and } s>(\sqrt{s}-\sqrt{r})^2+\beta.\end{cases}$$

Therefore,

$$\begin{aligned}\Delta_{1,0}(s;r,\beta) &= \frac{s^{1/4}}{2(\pi\log p)^{1/4}(\sqrt{s}-\sqrt{r})}p^{1/2-\beta-(\sqrt{s}-\sqrt{r})^2+s/2}I(r<s)\\ &\quad+(s\pi\log p)^{1/4}(r/s)^{1/4}p^{1/2-\beta+s/2}I(r>s).\end{aligned}\tag{A.24}$$

Replicating the above proof for the L_2 test, it can be shown that, for $\gamma=0$ and 1,

$$\hat{\mathcal{M}}_{\gamma n}\sim\max_{s\in\mathcal{S}_n}\Delta_{\gamma,0}(s;r,\beta).$$

At last, we will compare $\max_{s\in\mathcal{S}_n}\Delta_{\gamma,0}(s;r,\beta)$ for $\gamma=0,1$ and 2 when $r>2\beta-1$. Let $s_n^*=\arg\max\{s:s\in\mathcal{S}_n\cap(2\beta-1,r)\}$ be a threshold in $(2\beta-1,r)$ that is closest to r. Then the maximal value of $\Delta_{\gamma,0}(s,r,\beta)$ over $\mathcal{S}_n$ is attained at s_n^*. Note that such s_n^* exists with probability 1. To show this point, it is enough to show that $\mathcal{S}_n\cap(2\beta-1,r)\neq\varnothing$, which is equivalent to showing that $P(\bigcup_{i=1}^p\{Y_{i,n}\in((4\beta-2)\log p,2r\log p)\})\to 1$. Let $\{k_1,\dots,k_q\}\in(1,\dots,p)$ be a sub-sequence such that $q\to\infty$ and $k_{\min}=\min_j|k_j-k_{j-1}|\to\infty$. Let $D_n=\prod_{i=k_1}^{k_q}P(\{Y_{i,n}\in((4\beta-2)\log p,2r\log p)^c\})-P(\bigcap_{i=k_1}^{k_q}\{Y_{i,n}\in((4\beta-2)\log p,2r\log p)^c\})$. By mixing assumption (C.5) and the triangle inequality, it can be seen that $|D_n|\le q\alpha_Z(k_{\min})\to 0$ as $n\to\infty$. Then it follows that

$$
\begin{aligned}
&P\left(\bigcup_{i=1}^p\{Y_{i,n}\in\big((4\beta-2)\log p,2r\log p\big)\}\right)\\
&\qquad\ge P\left(\bigcup_{i=k_1}^{k_q}\{Y_{i,n}\in\big((4\beta-2)\log p,2r\log p\big)\}\right)\\
&\qquad=1-P\left(\bigcap_{i=k_1}^{k_q}\{Y_{i,n}\in\big((4\beta-2)\log p,2r\log p\big)^c\}\right)\\
&\qquad=1-\prod_{i=k_1}^{k_q}P\big(\{Y_{i,n}\in\big((4\beta-2)\log p,2r\log p\big)^c\}\big)+D_n\to 1,
\end{aligned}
$$

where we used $P(\{Y_{i,n}\in((4\beta-2)\log p,2r\log p)^c\})<1$ for all $i=1,\dots,p$. Comparing (A.8), (A.22) and (A.24), we see that $\Delta_{0,0}(s_n^*;r,\beta)<\Delta_{1,0}(s_n^*;r,\beta)<\Delta_{2,0}(s_n^*;r,\beta)$.

It follows that, for $r>2\beta-1$,

$$\max_{s\in\mathcal{S}_n}\Delta_{0,0}(s;r,\beta)<\max_{s\in\mathcal{S}_n}\Delta_{1,0}(s;r,\beta)<\max_{s\in\mathcal{S}_n}\Delta_{2,0}(s;r,\beta).$$

Therefore, asymptotically with probability 1, $\hat{\mathcal{M}}_{0n}<\hat{\mathcal{M}}_{1n}<\hat{\mathcal{M}}_{2n}$, which results in $\Omega_0(r,\beta)\le\Omega_1(r,\beta)\le\Omega_2(r,\beta)$. This completes the proof. □

Acknowledgments. The authors thank the Editor, an Associate Editor and two referees for insightful and constructive comments which have improved the presentation of the paper. We are also very grateful to Dr. Jiashun Jin and Dr. Cun-Hui Zhang for stimulating discussions.

SUPPLEMENTARY MATERIAL

A supplement to "Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence" (DOI: 10.1214/13-AOS1168SUPP; .pdf). The supplementary material contains proofs for Proposition 1 and Theorem 1 in Section 2.

REFERENCES

ANDREWS, D. and POLLARD, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Int. Statist. Rev.* **62** 119–132.

ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2012a). Detection of correlations. *Ann. Statist.* **40** 412–435. MR3014312

ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2012b). Detecting positive correlations in a multivariate sample. Available at arXiv:1202.5536v1 [math.ST].

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042

CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 629–662. MR2867452

CAI, T. and WU, Y. (2012). Optimal detection for sparse mixtures. Unpublished manuscript.

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

DELAIGLE, A. and HALL, P. (2009). Higher criticism in the context of unknown distribution, nonindependence and classification. In *Perspectives in Mathematical Sciences. I. Stat. Sci. Interdiscip. Res.* **7** 109–138. World Sci. Publ., Hackensack, NJ. MR2581742

DELAIGLE, A., HALL, P. and JIN, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 283–301. MR2815777

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

DONOHO, D. and JIN, J. (2008). Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proc. Natl. Acad. Sci. USA* **105** 14790–14795.

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089

DOUKHAN, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**. Springer, New York. MR1312160

FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674–688. MR1395735

HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. MR2387976

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357

INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. MR1456646

JING, B.-Y., SHAO, Q.-M. and ZHOU, W. (2008). Towards a universal self-normalized moderate deviation. *Trans. Amer. Math. Soc.* **360** 4263–4285. MR2395172

JOE, H. (1997). *Multivariate Models and Dependence Concepts. Monographs on Statistics and Applied Probability* **73**. Chapman & Hall, London. MR1462613

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York. MR0691492

PETROV, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables. Oxford Studies in Probability* **4**. Oxford Univ. Press, New York. MR1353441

PISIER, G. (1983). Some applications of the metric entropy condition to harmonic analysis. In *Banach Spaces, Harmonic Analysis, and Probability Theory* (*Storrs, Conn.*, 1980/1981). *Lecture Notes in Math.* **995** 123–154. Springer, Berlin. MR0717231

SHAO, Q.-M. (1997). Self-normalized large deviations. *Ann. Probab.* **25** 285–328. MR1428510

SIBUYA, M. (1960). Bivariate extreme statistics. I. *Ann. Inst. Statist. Math. Tokyo* **11** 195–210. MR0115241

TUKEY, J. W. (1976). T13 N: The higher criticism. Course Notes, Statistics 411, Princeton Univ.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. MR1385671

WANG, Q. and HALL, P. (2009). Relative errors in central limit theorems for Student's t statistic, with applications. *Statist. Sinica* **19** 343–354. MR2487894

ZHONG, P. S., CHEN, S. X. and XU, M. (2013). Supplement to "Tests alternative to higher criticism for high dimensional means under sparsity and column-wise dependence." DOI:10.1214/13-AOS1168SUPP.

P.-S. ZHONG
DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824
USA
E-MAIL: pszhong@stt.msu.edu

S. X. CHEN
DEPARTMENT OF BUSINESS STATISTICS AND ECONOMETRICS
GUANGHUA SCHOOL OF MANAGEMENT AND CENTER FOR STATISTICAL SCIENCE
PEKING UNIVERSITY
BEIJING 100871
CHINA
AND
DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50011-1210
USA
E-MAIL: songchen@iastate.edu

M. XU
DEPARTMENT OF BUSINESS STATISTICS AND ECONOMETRICS
GUANGHUA SCHOOL OF MANAGEMENT AND CENTER FOR STATISTICAL SCIENCE
PEKING UNIVERSITY
BEIJING 100871
CHINA
E-MAIL: minyaxu@gsm.pku.edu.cn