cemmap

centre for microdata methods and practice

Testing many moment inequalities

Victor Chernozhukov
Denis Chetverikov
Kengo Kato

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP65/13

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

TESTING MANY MOMENT INEQUALITIES

VICTOR CHERNOZHUKOV, DENIS CHETVERIKOV, AND KENGO KATO

Abstract. This paper considers the problem of testing many moment inequalities where the number of moment inequalities, denoted by p, is possibly much larger than the sample size n. There are variety of economic applications where the problem of testing many moment inequalities appears; a notable example is the entry model of Ciliberto and Tamer (2009) where $p = 2^{m+1}$ with m being the number of firms. We consider the test statistic given by the maximum of p Studentized (or t-type) statistics, and analyze various ways to compute critical values for the test. Specifically, we consider critical values based upon (i) the union (Bonferroni) bound combined with a moderate deviation inequality for self-normalized sums, (ii) the multiplier bootstrap. We also consider two step variants of (i) and (ii) by incorporating moment selection. We prove validity of these methods, showing that under mild conditions, they lead to tests with error in size decreasing polynomially in n while allowing for p being much larger than n; indeed p can be of order $\exp(n^c)$ for some $c > 0$. Importantly, all these results hold without any restriction on correlation structure between p Studentized statistics, and also hold uniformly with respect to suitably wide classes of underlying distributions. We also show that all the tests developed in this paper are asymptotically minimax optimal when p grows with n.

1. Introduction

In this paper, we are interested in testing *many* moment inequalities where the number of moment inequalities, denoted by p, is possibly much larger than the sample size n. Let $X_1, \dots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random vectors in $\mathbb{R}^p$, where $X_i = (X_{i1}, \dots, X_{ip})^T$. For $1 \leq j \leq p$, write $\mu_j := \mathrm{E}[X_{1j}]$. We are interested in testing the null hypothesis

$$H_0 : \mu_j \leq 0,\ 1 \leq \forall j \leq p, \tag{1}$$

Date: First version: April 2013. This version: December 30, 2013.

Key words and phrases. Bonferroni bound, many moment inequalities, moderate deviation, multiplier bootstrap, non-asymptotic bound, self-normalized sum.

We are grateful to Adam Rosen, Azeem Shaikh, participants at Cowles Summer Conference 2013, Asian Meeting of Econometric Society 2013, Bernoulli Society Satellite Conference, and seminar participants at UCL. V. Chernozhukov and D. Chetverikov are supported by a National Science Foundation grant. K. Kato is supported by the Grant-in-Aid for Young Scientists (B) (25780152), the Japan Society for the Promotion of Science.

against the alternative

$$H_1 : \mu_j > 0, \ 1 \leq \exists j \leq p. \tag{2}$$

We refer to (1) as the moment inequalities, and we say that the jth moment inequality is satisfied (violated) if $\mu_j \leq 0$ ($\mu_j > 0$). Thus H_0 is the hypothesis that all the moment inequalities are satisfied.

There are variety of economic applications where the problem of testing many moment inequalities appears. One example is the model where a consumer is selecting a bundle of products for purchase and moment inequalities come from a revealed preference argument (see Pakes, 2010). In this example, one typically has *many* moment inequalities because the number of different combinations of products from which the consumer is selecting is huge. Another example is a market structure model of Ciliberto and Tamer (2009) where the number of moment inequalities equals the number of possible combinations of firms presented in the market, which is exponentially large in the number of firms that could potentially enter the market. Another example is a dynamic model of imperfect competition of Bajari, Benkard, and Levin (2007), where deviations from optimal policy serve to define many moment inequalities. Other prominent examples leading to many moment inequalities are developed in Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011), and Chesher, Rosen, and Smolinski (2013), where moment inequalities are used to provide sharp identification regions for parameters in partially identified models. Many examples above have a very important feature – the large number of inequalities generated are "unstructured" in the sense that they can not be viewed as some unconditional moment inequalities generated from a small number of conditional inequalities with a low-dimensional conditioning variable.[1] This means that the existing methods for conditional moment inequalities, albeit fruitful in many cases, do not address this type of framework, and our methods are precisely aimed at dealing with this important case. We thus view our methods as strongly complementary to the existing literature.

We consider the test statistic given by the maximum over p Studentized (or t-type) statistics (see (4) ahead for the formal definition), and propose a number of methods for computing critical values. Specifically, we consider critical values based upon (i) the union (Bonferroni) bound combined with a moderate deviation inequality for self-normalized sums, and (ii) the multiplier bootstrap. We will call the first option the *SN method* (SN refers to the abbreviation of "Self-Normalized"), and the second option the *MB method* (MB refers to the abbreviation of "Multiplier Bootstrap"). The SN

[1]A small number of conditional inequalities gives rise to a large number of unconditional inequalities, but these have certain continuity and tightness structure, which the literature on conditional moment inequalities heavily exploits/relies upon. Our approach does not exploit/rely upon such structure and can handle both many unstructured moment inequalities as well as many structured moment inequalities arising from conversion of a small number of conditional inequalities.

method is analytical and hence easier to implement, while the MB method is simulation-based and computationally harder, but leads to less conservative tests as it is able to account for correlation between p Studentized statistics. We also consider two-step methods by incorporating moment selection procedures using either the SN or MB methods.

We prove validity of these methods for computing critical values, uniformly in suitable classes of common distributions of X_i. We derive non-asymptotic bounds on the rejection probabilities, where the qualification "non-asymptotic" means that the bounds hold with fixed n (and p, and all the other parameters), and the dependence of the constants involved in the bounds are stated explicitly. Notably, under mild conditions, these methods lead to the error in size decreasing polynomially in n, while allowing for p much larger than n, that is, typically, allowing for p proportional to $\exp(n^c)$ for some $c > 0$. We also show that all the tests developed in this paper are asymptotically minimax optimal when $p = p_n \to \infty$ as $n \to \infty$.

The problem of testing moment inequalities described above is a "dual" of that of constructing confidence regions for identifiable parameters in partially identified models where identified sets are given by (unconditional) moment inequalities, in the sense that any test of size (approximately) $\alpha \in (0, 1)$ for the former problem will lead to a confidence region for the latter problem with coverage (approximately) at least $1 - \alpha$ (see Romano and Shaikh, 2008). Therefore, our results on testing moment inequalities are immediately transferred to those on construction of confidence regions for identifiable parameters in partially identified models. That is, our methods for computing critical values lead to methods of construction of confidence regions with coverage error decreasing polynomially in n while allowing for $p \gg n$. Importantly, these coverage results hold uniformly in suitably wide classes of underlying distributions, so that the resulting confidence regions are *(asymptotically) honest* to such classes (see Section 5 for the precise meaning).

The literature on testing (unconditional) moment inequalities is large; see, for example, Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Canay (2010), Bugni (2011), Andrews and Jia-Barwick (2012), and Romano, Shaikh, and Wolf (2013). However, these papers deal only with a finite (and fixed) number of moment inequalities. There are also several papers on testing conditional moment inequalities, which can be treated as an infinite number of unconditional moment inequalities; see Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013), Lee, Song, and Whang (2013a,b), Armstrong (2011), Chetverikov (2011), and Armstrong and Chan (2012). However, when unconditional moment inequalities come from conditional ones, they inherit from original inequalities certain correlation structure that facilitates the analysis of such moment inequalities. In contrast, we are interested in treating many moment inequalities without assuming any correlation structure, motivated by important examples such as those in Cilberto and Tamer (2009) and Pakes (2010). Menzel (2008) considered estimation of identified sets when there are many moment inequalities using GMM criterion functions, thereby extending Chernozhukov, Hong, and Tamer (2007) to the case where the number of moment inequalities grows as the sample size increases. His results are, however, restricted to the case where p grows at most of cubic root of n (and hence p must be much smaller than n), and his approach and test statistics are quite different from ours.

The problem of testing many moment inequalities is related to that of multiple hypothesis testing. The difference is that the emphasis in multiple hypothesis testing is on improving power given that some inequalities are not satisfied, and the emphasis in testing many moment inequalities is on improving power given that some inequalities are not binding.

The remainder of the paper is organized as follows. In Section 2, we build our test statistic. In Section 3, we derive various ways of computing critical values, including the SN and MB methods and their two-step variants discussed above, for the test and prove their validity. In Section 4, we show asymptotic minimax optimality of our tests. In Section 5, we present the corresponding results on construction of confidence regions for identifiable parameters in partially identified models. In Section 6, we consider, as an extension, critical values for our test statistic based on the empirical (or Efron's) bootstrap. All the technical proofs are deferred to the Appendix.

1.1. **Notation and convention.** We shall obey the following notation. For an arbitrary sequence $\{z_i\}_{i=1}^n$, we write $\mathbb{E}_n[z_i] = n^{-1}\sum_{i=1}^n z_i$. For $a, b \in \mathbb{R}$, we use the notation $a \vee b = \max\{a, b\}$. For any finite set J, let $|J|$ denote the number of elements in J. The transpose of a vector z is denoted by z^T. Moreover, we use the notation $X_1^n = \{X_1, \dots, X_n\}$. In this paper, we (implicitly) assume that the quantities such as $X_1, \dots, X_n$ and p are all indexed by n. We are primarily interested in the case where $p = p_n \to \infty$ as $n \to \infty$. However, in most cases, we suppress the dependence of these quantities on n for the notational convenience. Finally, throughout the paper, we assume that $n \geq 2$ and $p \geq 2$.

2. Test statistic

We begin with preparing some notation. Recall that $\mu_j = \mathrm{E}[X_{1j}]$. We assume that

$$\mathrm{E}[X_{1j}^2] < \infty,\ \sigma_j^2 := \mathrm{Var}(X_{1j}) > 0,\ 1 \leq \forall j \leq p. \tag{3}$$

For $1 \leq j \leq p$, let $\widehat{\mu}_j$ and $\widehat{\sigma}_j^2$ denote the sample mean and variance of $X_{1j}, \dots, X_{nj}$, respectively, that is,

$$\widehat{\mu}_j = \mathbb{E}_n[X_{ij}] = \frac{1}{n}\sum_{i=1}^n X_{ij},\ \widehat{\sigma}_j^2 = \mathbb{E}_n[(X_{ij} - \mathbb{E}_n[X_{ij}])^2] = \frac{1}{n}\sum_{i=1}^n (X_{ij} - \widehat{\mu}_j)^2.$$

Alternatively, we can use $\tilde{\sigma}_j^2 = (1/(n-1))\sum_{i=1}^n (X_{ij} - \widehat{\mu}_j)^2$ instead of $\widehat{\sigma}_j^2$, which does not alter the overall conclusions of the theorems ahead. In all what follows, we will use $\widehat{\sigma}_j^2$.

There are several different statistics that can be used for testing the null hypothesis (1). Among all possible statistics, it is natural to consider statistics that take large values when some of $\widehat{\mu}_j$ are large. In this paper, we focus on the statistic that takes large values when at least one of $\widehat{\mu}_j$ is large. One can also consider either non-Studentized or Studentized versions of the test statistic. For a non-Studentized statistic, we mean a function of $\widehat{\mu}_1, \dots, \widehat{\mu}_p$, and for a Studentized statistic, we mean a function of $\widehat{\mu}_1/\widehat{\sigma}_1, \dots, \widehat{\mu}_p/\widehat{\sigma}_p$. Studentized statistics are often considered preferable. In particular, they are scale-invariant (that is, multiplying $X_{1j}, \dots, X_{nj}$ by a scalar value does not change the value of the test statistic), and they typically spread power evenly among different moment inequalities $\mu_j \le 0$. See Romano and Wolf (2005) for a detailed comparison of Studentized versus non-Studentized statistics in a related context of multiple hypothesis testing. In our case, Studentization also has an advantage that it allows us to derive an analytical critical value for the test under weak moment conditions. In particular, for our SN critical values, we will only require finiteness (existence) of $\mathrm{E}[|X_{1j}|^q]$ for some $2 < q \le 3$ (see Section 3.1.1). As far as multiplier bootstrap critical values are concerned, our theory can cover a non-Studentized statistic but Studentization leads to easily interpretable regularity conditions. For these reasons, in this paper we study the Studentized version of the test statistic.

To be specific, we consider and focus on the following test statistic:

$$T = \max_{1 \le j \le p} \frac{\sqrt{n}\widehat{\mu}_j}{\widehat{\sigma}_j}. \tag{4}$$

Large values of T indicate that H_0 is likely to be violated, so that it would be natural to consider the test of the form

$$T > c \Rightarrow \text{reject } H_0, \tag{5}$$

where c is a critical value suitably chosen in such a way that the test has approximately size $\alpha \in (0,1)$. We will consider various ways for calculating critical values and prove their validity.

Rigorously speaking, the test statistic T is not defined when $\widehat{\sigma}_j^2 = 0$ for some $1 \le j \le p$. In such cases, we interpret the meaning of "$T > c$" in (5) as

$$\sqrt{n}\widehat{\mu}_j > c\widehat{\sigma}_j, \ 1 \le \exists j \le p,$$

which makes sense even if $\widehat{\sigma}_j^2 = 0$ for some $1 \le j \le p$. We will obey such conventions if necessary without further mentioning.

Other types of test statistics are possible. For example, one alternative is the test statistic of the form

$$T' = \sum_{j=1}^p \left(\max\{\widehat{\mu}_j/\widehat{\sigma}_j, 0\}\right)^2.$$

The statistic T' has an advantage that it is less sensitive to outliers. However, T' leads to good power only if many moments are violated simultaneously. In general, T' is preferable against T if the researcher is interested in detecting deviations when many inequalities are violated simultaneously, and T is preferable against T' if the main interest is in detecting deviations when at least one moment inequality is violated too much.

Another alternative is the test statistic of the form

$$T'' = \min_{t \leq 0} (\widehat{\mu} - t)^T \widehat{\Sigma}^{-1} (\widehat{\mu} - t),$$

where $\widehat{\mu} = (\widehat{\mu}_1, \dots, \widehat{\mu}_p)^T$, $t = (t_1, \dots, t_p)^T \leq 0$ means $t_j \leq 0$ for all $1 \leq j \leq p$, and $\widehat{\Sigma}$ is some p by p symmetric positive definite matrix. This statistic in the context of testing moment inequalities was first studied by Rosen (2008) when the number of moment inequalities p is fixed; see also Wolak (1991) for the analysis of this statistic in a different context. Typically, one wants to take $\widehat{\Sigma}$ as a suitable estimate of the covariance matrix of X_1, denoted by Σ. However, when p is larger than n, it is not possible to consistently estimate Σ without imposing some structure (such as sparsity) on it. Moreover, the results of Bai and Saranadasa (1996) suggest that the statistic T' or its variants may lead to higher power than T'' even when p is smaller than but close to n.

3. Critical values

In this section, we study several methods to compute critical values for the test statistic T so that under H_0, the probability of rejecting H_0 does not exceed size α approximately. The basic idea for construction of critical values for T lies in the fact that under H_0,

$$T \leq \max_{1 \leq j \leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j, \tag{6}$$

where the equality holds when all the moment inequalities are binding, that is, $\mu_j = 0$ for all $1 \leq j \leq p$. Hence in order to make the test to have size α, it is enough to choose the critical value as (a bound on) the $(1-\alpha)$-quantile of the distribution of $\max_{1 \leq j \leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j$. We consider two approaches to construct such critical values: Self-Normalized and Multiplier Bootstrap methods. For each of these methods, we also consider its two-step variant by incorporating moment selection.

We will use the following notation. Pick any $\alpha \in (0, 1/2)$. Let

$$Z_{ij} = (X_{ij} - \mu_j)/\sigma_j, \text{ and } Z_i = (Z_{ij}, \dots, Z_{ip})^T. \tag{7}$$

Observe that $\mathrm{E}[Z_{ij}] = 0$ and $\mathrm{E}[Z_{ij}^2] = 1$. Define

$$M_{n,k} = \max_{1 \leq j \leq p} (\mathrm{E}[|Z_{1j}|^k])^{1/k}, \ k = 3, 4, \quad B_n = (\mathrm{E}[\max_{1 \leq j \leq p} Z_{1j}^4])^{1/4}.$$

Note that $B_n \geq M_{n,4} \geq M_{n,3} \geq 1$.

3.1. **Self-Normalized methods.**

3.1.1. *One-step method.* The Self-Normalized method (abbreviated as SN method in what follows) we consider is based upon the union (Bonferroni) bound combined with a moderate deviation inequality for self-normalized sums. Because of inequality (6), under H_0,

$$\mathrm{P}(T > c) \le \sum_{j=1}^{p} \mathrm{P}(\sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j > c).$$

At a first sight, this bound is too crude when p is large since, as long as X_{ij} has polynomial tails, the value of c that makes the sum on the right-hand side of the equation above bounded by size α depends polynomially on p, which would make the test too conservative. However, we will exploit the self-normalizing nature of the quantity $\sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j$ so that the resulting critical value depends on p only through its logarithm. In addition, in spite of the fact that the SN method is based on the union (Bonferroni) bound, we will show in Section 4 that the resulting test in asymptotically minimax optimal when $p = p_n \to \infty$ as $n \to \infty$.

For $1 \le j \le p$, define

$$U_j = \sqrt{n}\mathbb{E}_n[Z_{ij}]/\sqrt{\mathbb{E}_n[Z_{ij}^2]}.$$

By a simple algebra, we see that

$$\sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j = U_j/\sqrt{1 - U_j^2/n},$$

where the right side is increasing in U_j as long as $U_j \ge 0$. Hence under H_0,

$$\mathrm{P}(T > c) \le \sum_{j=1}^{p} \mathrm{P}\left(U_j > c/\sqrt{1 + c^2/n}\right), \ c \ge 0. \tag{8}$$

Because, from the moderate deviation inequality for self-normalized sums of Jing, Shao, and Wang (2003) (see Lemma A.1 in the Appendix), for moderately large $c \ge 0$,

$$\mathrm{P}\left(U_j > c/\sqrt{1 + c^2/n}\right) \approx \mathrm{P}(N(0,1) > c/\sqrt{1 + c^2/n}),$$

we consider to take the critical value as

$$c^{SN}(\alpha) = \frac{\Phi^{-1}(1 - \alpha/p)}{\sqrt{1 - \Phi^{-1}(1 - \alpha/p)^2/n}}, \tag{9}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution, and $\Phi^{-1}(\cdot)$ is its quantile function. We will call $c^{SN}(\alpha)$ the (one-step) SN critical value with size α as its derivation depends on the moderate deviation inequality for self-normalized sums. Note that

$$\Phi^{-1}(1 - \alpha/p) \sim \sqrt{\log(p/\alpha)},$$

so that $c^{SN}(\alpha)$ depends on p only through $\log p$.

The following theorem provides a non-asymptotic bound on the probability that the test statistic T exceeds the SN critical value $c^{SN}(\alpha)$ under H_0

and shows that the bound converges to α under mild regularity conditions, thereby validating the SN method.

Theorem 3.1 (Validity of one-step SN method)**.** *Suppose that* $\Phi^{-1}(1-\alpha/p) \leq n^{1/6}/M_{n,3}$*. Then under* H_0*,*

$$P(T > c^{SN}(\alpha)) \leq \alpha \left[1 + Kn^{-1/2}M_{n,3}^3\{1+\Phi^{-1}(1-\alpha/p)\}^3\right], \qquad (10)$$

where K *is a universal constant. Hence if there exist constants* $0 < c_1 < 1/2$ *and* $C_1 > 0$ *such that*

$$M_{n,3}^3 \log^{3/2}(p/\alpha) \leq C_1 n^{1/2-c_1}, \qquad (11)$$

then under H_0*,*

$$P(T > c^{SN}(\alpha)) \leq \alpha + Cn^{-c_1},$$

where C *is a constant depending only on* c_1, C_1*. Moreover, this bound holds uniformly with respect to the common distribution of* X_i *for which (3) and (11) are verified.*

Comment 3.1. The theorem assumes that $\max_{1\leq j\leq p} E[|X_{1j}|^3] < \infty$ but allows this quantity to diverges as $n \to \infty$ (recall $p = p_n$). In principle, $M_{n,3}^3$ that appears in the theorem can be replaced by $\max_{1\leq j\leq p} E[|X_{1j}|^{2+\nu}]$ for $0 < \nu \leq 1$, which would further weaken moment conditions; however, for the sake of simplicity of presentation, we do not explore this generalization.

3.1.2. *Two-step method.* We now turn to combine the SN method with moment selection. We begin with stating the motivation for moment selection.

Observe that when $\mu_j < 0$ for some $1 \leq j \leq p$, inequality (6) becomes strict, so that when there are many j for which μ_j are small, the resulting test with one-step SN critical values would tend to be more conservative. Hence it is intuitively clear that, in order to improve the power of the test, it is better to exclude j for which μ_j are below some (negative) threshold when computing critical values. This is the basic idea behind moment selection.

More formally, let $0 < \beta_n < \alpha/2$ be some constant. For generality, we allow β_n to depend on n; in particular, β_n is allowed to decrease to zero as the sample size n increases. Let $c^{SN}(\beta_n)$ be the SN critical value with size β_n, and define the set $\widehat{J}_{SN} \subset \{1,\dots,p\}$ by

$$\widehat{J}_{SN} = \{j \in \{1,\dots,p\} : \sqrt{n}\widehat{\mu}_j > -2\widehat{\sigma}_j c^{SN}(\beta_n)\}. \qquad (12)$$

Let $\widehat{k}$ denote the number of elements in $\widehat{J}_{SN}$, that is,

$$\widehat{k} = |\widehat{J}_{SN}|.$$

Then the two-step SN critical value is defined by

$$c^{SN,MS}(\alpha) = \begin{cases} \frac{\Phi^{-1}(1-(\alpha-2\beta_n)/\widehat{k})}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta_n)/\widehat{k})^2/n}}, & \text{if } \widehat{k} \geq 1, \\ 0, & \text{if } \widehat{k} = 0. \end{cases} \qquad (13)$$

The following theorem establishes validity of this critical value.

Theorem 3.2 (Validity of two-step SN method)**.** *Suppose that* $\sup_{n\geq 1}\beta_n \leq \alpha/3$ *and there exist constants* $0 < c_1 < 1/2$ *and* $C_1 > 0$ *such that*

$$M_{n,3}^3 \log^{3/2}(p/\beta_n) \leq C_1 n^{1/2-c_1}, \text{ and } B_n^2 \log^2(p/\beta_n) \leq C_1 n^{1/2-c_1}. \quad (14)$$

Then there exist positive constants c, C *depending only on* c_1, C_1 *such that under* H_0,

$$\mathrm{P}(T > c^{SN,MS}(\alpha)) \leq \alpha + Cn^{-c}.$$

Moreover, this bound holds uniformly with respect to the common distribution of X_i *for which (3) and (14) are verified.*

3.2. **Multiplier Bootstrap (MB) methods.** In this section, we consider the Multiplier Bootstrap (MB) for calculating critical values. The methods considered in this section are computationally harder than those in the previous section but they lead to less conservative tests. In particular, we will show that when all the moment inequalities are binding (that is, $\mu_j = 0$ for all $1 \leq j \leq p$), the asymptotic size of the tests based on these methods coincides with the nominal size.

3.2.1. *One-step method.* We first consider the one-step method. Recall that, in order to make the test to have size α, it is enough to choose the critical value as (a bound on) the $(1-\alpha)$-quantile of the distribution of

$$\max_{1\leq j\leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j.$$

The SN method finds such a bound by using the union (Bonferroni) bound and the moderate deviation inequality for self-normalized sums. However, the SN method may be conservative as it ignores correlation between the coordinates in X_i.

Alternatively, we consider here a Gaussian approximation. Observe first that under suitable regularity conditions,

$$\max_{1\leq j\leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j \approx \max_{1\leq j\leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\sigma_j = \max_{1\leq j\leq n} \sqrt{n}\mathbb{E}_n[Z_{ij}],$$

where $Z_i = (Z_{i1}, \ldots, Z_{ip})^T$ are defined in (7). *When* p *is fixed*, the central limit theorem guarantees that as $n \to \infty$,

$$\sqrt{n}\mathbb{E}_n[Z_i] \stackrel{d}{\to} Y, \text{ with } Y = (Y_1, \ldots, Y_p)^T \sim N(0, \mathrm{E}[Z_1 Z_1^T]),$$

which, by the continuous mapping theorem, implies that

$$\max_{1\leq j\leq p} \sqrt{n}\mathbb{E}_n[Z_{ij}] \stackrel{d}{\to} \max_{1\leq j\leq p} Y_j.$$

Hence in this case it is enough to take the critical value as the $(1-\alpha)$-quantile of the distribution of $\max_{1\leq j\leq p} Y_j$.

When p *grows with* n, however, the concept of convergence in distribution does not apply, and different tools should be used to derive an appropriate critical value for the test. One possible approach is to use a Berry-Esseen theorem that provides a (suitable) non-asymptotic bound between the distributions of $\sqrt{n}\mathbb{E}_n[Z_i]$ and Y; see, for example, Götze (1991) and Bentkus

(2003). However, such Berry-Esseen bounds require p to be small in comparison with n in order to guarantee that the distribution of $\sqrt{n}\mathbb{E}_n[Z_i]$ is close to that of Y. Another possible approach is to compare the distributions of $\max_{1\le j\le p}\sqrt{n}\mathbb{E}_n[Z_{ij}]$ and $\max_{1\le j\le p} Y_j$ directly, avoiding the comparison of distributions of the whole vectors $\sqrt{n}\mathbb{E}_n[Z_i]$ and Y. Our recent work (Chernozhukov, Chetverikov, and Kato, 2013a) shows that, under mild regularity conditions, the distribution of $\max_{1\le j\le p}\sqrt{n}\mathbb{E}_n[Z_{ij}]$ can be approximated by that of $\max_{1\le j\le p} Y_j$ in the sense of Kolmogorov distance *even when p is larger or even much larger than n.*[2] This result implies that we can still use the $(1-\alpha)$-quantile of the distribution of $\max_{1\le j\le p} Y_j$ even when p grows with n and is potentially much larger than n.[3]

Still, the distribution of $\max_{1\le j\le p} Y_j$ is typically unknown because the covariance structure of Y is unknown. Hence we will approximate the distribution of $\max_{1\le j\le p} Y_j$ by the following multiplier bootstrap procedure:

Algorithm (Multiplier bootstrap)**.**

1. Generate independent standard normal random variables $\epsilon_1,\dots,\epsilon_n$ independent of the data $X_1^n=\{X_1,\dots,X_1^n\}$.

2. Construct the multiplier bootstrap test statistic
$$W=\max_{1\le j\le p}\frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_j)]}{\widehat{\sigma}_j}. \tag{15}$$

3. Calculate $c^{MB}(\alpha)$ as
$$c^{MB}(\alpha)=\text{conditional } (1-\alpha)\text{-quantile of } W \text{ given } X_1^n. \tag{16}$$

We will call $c^{SN}(\alpha)$ the (one-step) Multiplier Bootstrap (MB) critical value with size α. In practice conditional quantiles of W can be computed with any precision by using simulation.

Intuitively, it is expected that the multiplier bootstrap works well since conditional on the data X_1^n, the vector
$$\left(\frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_j)]}{\widehat{\sigma}_j}\right)_{1\le j\le p}$$
has the centered normal distribution with covariance matrix
$$\mathbb{E}_n\left[\frac{(X_{ij}-\widehat{\mu}_j)}{\widehat{\sigma}_j}\frac{(X_{ik}-\widehat{\mu}_k)}{\widehat{\sigma}_k}\right],\ 1\le j,k\le p,$$
which should be close to the covariance matrix of the vector Y. Indeed, by Theorem 2 in Chernozhukov, Chetverikov, and Kato (2013b), the primary

[2]The Kolmogorov distance between the distributions of two random variables ξ and η is defined by $\sup_{t\in\mathbb{R}}|\mathrm{P}(\xi\le t)-\mathrm{P}(\eta\le t)|$.

[3]Some applications of this result can be found in Chetverikov (2011, 2012), Wasserman, Kolar and Rinaldo (2013), and Chazal, Fasy, Lecci, Rinaldo, and Wasserman (2013).

factor for the bound on the Kolmogorov distance between the conditional distribution of W and the distribution of $\max_{1\leq j\leq p} Y_j$ is

$$\max_{1\leq j,k\leq p} \left| \mathbb{E}_n \left[\frac{(X_{ij}-\widehat{\mu}_j)}{\widehat{\sigma}_j} \frac{(X_{ik}-\widehat{\mu}_k)}{\widehat{\sigma}_k} \right] - \mathrm{E}[Z_{1j}Z_{1k}] \right|,$$

which can be small even when $p \gg n$. The following theorem formally establishes validity of the MB critical value.

Theorem 3.3 (Validity of one-step MB method)**.** *Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$(M_{n,3}^3 \vee M_{n,4}^2 \vee B_n)^2 \log^{7/2}(pn) \leq C_1 n^{1/2-c_1}. \tag{17}$$

Then there exist positive constants c, C depending only on c_1, C_1 such that under H_0,

$$\mathrm{P}(T > c^{MB}(\alpha)) \leq \alpha + Cn^{-c}. \tag{18}$$

If $\mu_j = 0$ for all $1 \leq j \leq p$, then

$$|\mathrm{P}(T > c^{MB}(\alpha)) - \alpha| \leq Cn^{-c}. \tag{19}$$

Moreover, all these bounds hold uniformly with respect to the common distribution of X_i for which (3) and (17) are verified.

3.2.2. *Two-step method.* We now consider to combine the MB method with moment selection. To describe this procedure, let $0 < \beta_n < \alpha/2$ be some constant. As in the previous section, we allow β_n to depend on n. Let $c^{MB}(\beta_n)$ be the (one-step) MB critical value with size β_n. Define the set $\widehat{J}_{MB}$ by

$$\widehat{J}_{MB} = \{j \in \{1,\dots,p\} : \sqrt{n}\widehat{\mu}_j > -2\widehat{\sigma}_j c^{MB}(\beta_n)\}.$$

Then the two-step MB critical value $c^{MB,MS}(\alpha)$ is defined by the following procedure:

Algorithm (Multiplier bootstrap with moment selection)**.**

1. Generate independent standard normal random variables $\epsilon_1, \dots, \epsilon_n$ independent of the data X_1^n.
2. Construct the bootstrap test statistic

$$W_{\widehat{J}_{MB}} = \begin{cases} \max_{j\in\widehat{J}_{MB}} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_j)]}{\widehat{\sigma}_j}, & \text{if } \widehat{J}_{MB} \text{ is not empty,} \\ 0 & \text{if } \widehat{J}_{MB} \text{ is empty.} \end{cases}$$

3. Calculate $c^{MB,MS}(\alpha)$ as

$$c^{MB,MS}(\alpha) = \text{conditional } (1-\alpha+2\beta_n)\text{-quantile of } W_{\widehat{J}_{MB}} \text{ given } X_1^n. \tag{20}$$

The following theorem establishes validity of the two-step MB critical value.

Theorem 3.4 (Validity of two-step MB method)**.** *Suppose that the assumption of Theorem 3.3 is satisfied. Moreover, suppose that $\sup_{n\geq 1}\beta_n < \alpha/2$ and $\log(1/\beta_n) \leq C_1 \log n$. Then all the conclusions of Theorem 3.3 hold with $c^{MB}(\alpha)$ replaced by $c^{MB,MS}(\alpha)$.*

Comment 3.2. The selection procedure used in the theorem above is most closely related to those in Chernozhukov, Lee, and Rosen (2013) and in Chetverikov (2011). Other selection procedures were suggested in the literature in the framework when p is fixed. Specifically, Romano, Shaikh, and Wolf (2013) derived a moment selection method based on the construction of rectangular confidence sets for the vector $(\mu_1, \dots, \mu_p)^T$. To extend their method to high dimensional setting considered here, note that by (19), we have that $\mu_j \leq \widehat{\mu}_j + \widehat{\sigma}_j c^{MB}(\beta_n)/\sqrt{n}$ for all $1 \leq j \leq p$ with probability $1 - \beta_n$ asymptotically. Therefore, we can replace (6) with the following probabilistic inequality: under H_0,

$$\mathrm{P}\left(T \leq \max_{1 \leq j \leq p} \frac{\sqrt{n}(\widehat{\mu}_j - \mu_j + \tilde{\mu}_j)}{\widehat{\sigma}_j}\right) \geq 1 - \beta_n,$$

where

$$\tilde{\mu}_j = \min(\widehat{\mu}_j + \widehat{\sigma}_j c^{MB}(\beta_n)/\sqrt{n}, 0).$$

This suggests that we could obtain a critical value based on the distribution of bootstrap test statistic

$$\widehat{W} = \max_{1 \leq j \leq p} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \widehat{\mu}_j)] + \sqrt{n}\tilde{\mu}_j}{\widehat{\sigma}_j}.$$

For brevity, however, we leave analysis of this critical value for future research. □

3.3. **Hybrid method.** We have considered the one-step SN and MB methods and their two-step variants. In fact, we can consider "hybrids" of the SN and MB methods. For example, we can use the SN method for moment selection, and apply the MB method for the selected moment inequalities, which is a computationally more tractable alternative to the two-step MB method. For convenience of terminology, we will call this method the Hybrid (HB) method, which is formally described as follows: let $0 < \beta_n < \alpha/2$ be some constants, and recall the set $\widehat{J}_{SN} \subset \{1, \dots, p\}$ defined in (12).

Algorithm (Hybrid method)**.**

1. Generate independent standard normal random variables $\epsilon_1, \dots, \epsilon_n$ independent of the data X_1^n.
2. Construct the bootstrap test statistic

$$W_{\widehat{J}_{SN}} = \begin{cases} \max_{j \in \widehat{J}_{SN}} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \widehat{\mu}_j)]}{\widehat{\sigma}_j}, & \text{if } \widehat{J}_{SN} \text{ is not empty,} \\ 0 & \text{if } \widehat{J}_{SN} \text{ is empty.} \end{cases}$$

3. Calculate $c^{HB}(\alpha)$ as

$$c^{HB}(\alpha) = \text{conditional } (1 - \alpha + 2\beta_n)\text{-quantile of } W_{\widehat{J}_{SN}} \text{ given } X_1^n. \quad (21)$$

The following theorem establishes validity of the HB critical value $c^{HB}(\alpha)$.

Theorem 3.5 (Validity of HB critical value). *Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that (17) is verified. Moreover, suppose that $\sup_{n \geq 1} \beta_n \leq \alpha/3$ and $\log(1/\beta_n) \leq C_1 \log n$. Then all the conclusions of Theorem 3.3 hold with $c^{MB}(\alpha)$ replaced by $c^{HB}(\alpha)$.*

4. Minimax optimality

In this section, we show that the tests developed in this paper are asymptotically minimax optimal when $p = p_n \to \infty$ as $n \to \infty$. We start with deriving an upper bound on the power any procedure may have in testing (1) against (2).

Lemma 4.1 (Upper bounds on power). *Let $X_1, \dots, X_n \sim N(\mu, \Sigma)$ i.i.d. where $\Sigma = \operatorname{diag}\{\sigma_1^2, \dots, \sigma_p^2\}$ and $\sigma_j^2 > 0$ for all $1 \leq j \leq p$, and consider testing the null hypothesis $H_0 : \max_{1 \leq j \leq p} \mu_j \leq 0$ against the alternative $H_1 : \max_{1 \leq j \leq p}(\mu_j/\sigma_j) \geq \theta$ with $\theta > 0$ a constant. Denote by $\mathrm{E}_\mu[\cdot]$ the expectation under μ. Then for any test $\phi_n : (\mathbb{R}^p)^n \to [0,1]$ such that $\mathrm{E}_\mu[\phi_n(X_1, \dots, X_n)] \leq \alpha$ for all $\mu \in \mathbb{R}^p$ with $\max_{1 \leq j \leq p} \mu_j \leq 0$, we have*

$$\inf_{\max_{1 \leq j \leq p}(\mu_j/\sigma_j) \geq \theta} \mathrm{E}_\mu[\phi_n(X_1, \dots, X_n)] \leq \alpha + \mathrm{E}[|p^{-1} \textstyle\sum_{j=1}^p e^{\sqrt{n}\theta \xi_j - n\theta^2/2} - 1|], \tag{22}$$

where $\xi_1, \dots, \xi_p \sim N(0,1)$ i.i.d. Moreover if $p = p_n \to \infty$ as $n \to \infty$, we have

$$\lim_{n \to \infty} \mathrm{E}[|p_n^{-1} \textstyle\sum_{j=1}^{p_n} e^{\sqrt{n}\underline{\theta}_n \xi_j - n\underline{\theta}_n^2/2} - 1|] = 0,$$

where $\underline{\theta}_n = (1 - \epsilon_n)\sqrt{2(\log p_n)/n}$, and $\epsilon_n > 0$ is any sequence such that $\epsilon_n \to 0$ and $\epsilon_n \sqrt{\log p_n} \to \infty$ as $n \to \infty$.

Going back to the general setting described in Section 1, assume (3) and consider the test statistic T defined in (4). Pick any $\alpha \in (0, 1/2)$ and consider in general the test of the form

$$T > \widehat{c}(\alpha) \Rightarrow \text{reject } H_0,$$

where $\widehat{c}(\alpha)$ is a possibly data-dependent critical value which makes the test to have size approximately α.

Lemma 4.2 (Lower bounds on power). *Consider the setting described above. Suppose that there exists a constant $\underline{\epsilon} \geq 0$ possibly depending on α such that $\widehat{c}(\alpha) \leq (1 + \underline{\epsilon})\sqrt{2\log(p/\alpha)}$ with probability one. Then for every $\epsilon > 0$ and $\delta \in (0, 1/2)$, whenever*

$$\max_{1 \leq j \leq p} (\mu_j/\sigma_j) \geq \frac{(1-\delta)(1+\epsilon+\underline{\epsilon})}{1-2\delta} \sqrt{\frac{2\log(p/\alpha)}{n}},$$

we have

$$\mathrm{P}(T > \widehat{c}(\alpha)) \geq 1 - \frac{1}{2(1-\delta)^2 \epsilon^2 \log(p/\alpha)} - \mathrm{P}(\max_{1 \leq j \leq p} |\widehat{\sigma}_j/\sigma_j - 1| > \delta).$$

From this lemma, we have the following corollary:

Corollary 4.1 (Asymptotic minimax optimality)**.** *Let* $\widehat{c}(\alpha)$ *be any one of* $c^{SN}(\alpha)$*,* $c^{SN,MS}(\alpha)$*,* $c^{MB}(\alpha)$*,* $c^{MB,MS}(\alpha)$*, or* $c^{HB}(\alpha)$ *where we assume* $\sup_{n\geq 1}\beta_n \leq \alpha/3$ *whenever moment selection is used. Suppose there exist constants* $0 < c_1 < 1/2$ *and* $C_1 > 0$ *such that*

$$B_n^2 \log^{3/2} p \leq C_1 n^{1/2-c_1}. \tag{23}$$

Then there exist constants $c, C > 0$ *depending only on* α, c_1, C_1 *such that for every* $\epsilon \in (0,1)$*, whenever*

$$\max_{1\leq j\leq p}(\mu_j/\sigma_j) \geq (1+\epsilon+C\log^{-1/2} p)\sqrt{\frac{2\log(p/\alpha)}{n}},$$

we have

$$\mathrm{P}(T > \widehat{c}(\alpha)) \geq 1 - \frac{C}{\epsilon^2 \log(p/\alpha)} - Cn^{-c}.$$

Therefore when $p = p_n \to \infty$*, for any sequence* ϵ_n *satisfying* $\epsilon_n \to 0$ *and* $\epsilon_n\sqrt{\log p_n} \to \infty$*, as* $n \to \infty$*, we have (with keeping* α *fixed)*

$$\inf_{\max_{1\leq j\leq p}(\mu_j/\sigma_j)\geq\overline{\theta}_n} \mathrm{P}_\mu(T > \widehat{c}(\alpha)) \geq 1 - o(1), \tag{24}$$

where $\overline{\theta}_n = (1+\epsilon_n)\sqrt{2(\log p_n)/n}$ *and* P_μ *is the probability under* μ*. Moreover, the above asymptotic result (24) holds uniformly with respect to the sequence of common distributions of* X_i *for which (3) and (23) are verified with given* c_1, C_1*.*

Comparing the bounds in Lemma 4.1 and Corollary 4.1, we see that all the tests developed in this paper are asymptotically minimax optimal when $p = p_n \to \infty$ as $n \to \infty$ under mild regularity conditions.

5. Honest confidence regions for identifiable parameters in partially identified models

In this section, as in Romano and Shaikh (2008) and Romano, Shaikh, and Wolf (2013), we consider the related problem of constructing confidence regions for identifiable parameters in partially identified models.

Let $\xi_1,\dots,\xi_n$ be i.i.d. random variables taking values in a (generic) measurable space $(S,\mathcal{S})$ with common distribution P, let Θ be a parameter space which is a (Borel measurable) subset of a metric space (usually a Euclidean space), and let

$$g : S\times\Theta \to \mathbb{R}^p,\ (\xi,\theta)\mapsto g(\xi,\theta) = (g_1(\xi,\theta),\dots,g_p(\xi,\theta))^T,$$

be a jointly Borel measurable map. We consider the partially identified model where the identified set $\Theta_0(P)$ is given by

$$\Theta_0(P) = \{\theta\in\Theta : \mathrm{E}_P[g_j(\xi_1,\theta)] \leq 0,\ 1\leq \forall j \leq p\}.$$

Here E_P means that the expectation is taken with respect to P (similarly P_P means that the probability is taken with respect to P). We consider the problem of constructing confidence regions $\mathcal{C}_n(\alpha) = \mathcal{C}_n(\alpha; \xi_1, \dots, \xi_n) \subset \Theta$ such that for some constant $c, C > 0$, for all $n \geq 1$,

$$\inf_{P \in \mathcal{P}_n} \inf_{\theta \in \Theta_0(P)} \mathrm{P}_P(\theta \in \mathcal{C}_n(\alpha)) \geq 1 - \alpha - Cn^{-c}, \tag{25}$$

while allowing for $p > n$ (indeed we allow p to be much larger than n), where $0 < \alpha < 1$ and $\mathcal{P}_n$ is a suitable class of distributions on $(S, \mathcal{S})$.[4] We call confidence regions $\mathcal{C}_n(\alpha)$ for which (25) is verified *asymptotically honest to $\mathcal{P}_n$ with a polynomial rate*, where the term is inspired by Li (1989) and Chernozhukov, Chetverikov, and Kato (2013c).

We first state the required restriction of the class of distributions $\mathcal{P}_n$. We assume that for every $P \in \mathcal{P}_n$,

$$\Theta_0(P) \neq \emptyset, \text{ and } \mathrm{E}_P[g_j^2(\xi_1, \theta)] < \infty, \ \sigma_j^2(\theta, P) := \mathrm{Var}_P(g_j(\xi_1, \theta)) > 0, \quad 1 \leq \forall j \leq p, \ \forall \theta \in \Theta_0(P). \tag{26}$$

We construct confidence regions based upon duality between hypothesis testing and construction of confidence regions. For any given $\theta \in \Theta$, consider the statistic

$$T(\theta) = \max_{1 \leq j \leq p} \sqrt{n} \widehat{\mu}_j(\theta) / \widehat{\sigma}_j(\theta),$$

where

$$\widehat{\mu}_j(\theta) = \mathbb{E}_n[g_j(\xi_i, \theta)], \ \widehat{\sigma}_j^2(\theta) = \mathbb{E}_n[(g_j(\xi_i, \theta) - \widehat{\mu}_j(\theta))^2].$$

This statistic is a test statistic for the problem

$$H_\theta : \mu_j(\theta, P) \leq 0, \ 1 \leq \forall j \leq p,$$

against

$$H_\theta' : \mu_j(\theta, P) > 0, \ 1 \leq \exists j \leq p,$$

where $\mu_j(\theta, P) := \mathrm{E}_P[g_j(\xi_1, \theta)]$. Pick any $\alpha \in (0, 1/2)$. We consider the confidence region of the form

$$\mathcal{C}_n(\alpha) = \{\theta \in \Theta : T(\theta) \leq c(\alpha, \theta)\}, \tag{27}$$

where $c(\alpha, \theta)$ is a critical value such that $\mathcal{C}_n(\alpha)$ contains θ with probability (approximately) at least $1 - \alpha$ whenever $\theta \in \Theta_0(P)$.

Recall $c^{SN}(\alpha)$ defined in (9), and let

$$c^{SN,MS}(\alpha, \theta), c^{MB}(\alpha, \theta), c^{MB,MS}(\alpha, \theta), c^{HB}(\alpha),$$

be the two-step SN, one-step MB, two-step MB, and HB critical values defined in (13), (16), (20), and (21), respectively, with $X_i = (X_{i1}, \dots, X_{ip})^T$ replaced by $g(\xi_i, \theta) = (g_1(\xi_i, \theta), \dots, g_p(\xi_i, \theta))^T$. Moreover, let $\mathcal{C}_n^{SN}(\alpha)$ be the confidence region (27) with $c(\alpha, \theta) = c^{SN}(\alpha)$; define

$$\mathcal{C}_n^{SN,MS}(\alpha), \mathcal{C}_n^{MB}(\alpha), \mathcal{C}_n^{MB,MS}(\alpha), \mathcal{C}_n^{HB}(\alpha),$$

[4] We allow the class to depend on n, so that strictly speaking, we consider sequences of classes of distributions.

analogously. Finally, define

$$M_{n,k}(\theta,P) := \max_{1\leq j\leq p}(\mathrm{E}_P[|(g_j(\xi_1,\theta)-\mu_j(\theta,P))/\sigma_j(\theta,P)|^k])^{1/k},\ k=3,4,$$

$$B_n(\theta,P) := \left(\mathrm{E}_P\left[\max_{1\leq j\leq p}|(g_j(\xi_1,\theta)-\mu_j(\theta,P))/\sigma_j(\theta,P)|^4\right]\right)^{1/4}.$$

Let $0<c_1<1/2, C_1>0$ be given constants (we implicitly assume that C_1 is large enough). The following theorem is the main result of this section.

Theorem 5.1. *Let $\mathcal{P}_n^{SN}$ be the class of distributions P on $(S,\mathcal{S})$ for which (26) and (11) are verified with $M_{n,3}$ replaced by $M_{n,3}(\theta,P)$ for all $\theta\in\Theta_0(P)$; let $\mathcal{P}_n^{SN,MS}$ be the class of distributions P on $(S,\mathcal{S})$ for which (26) and (14) are verified with $M_{n,3}, B_n$ replaced by (respectively) $M_{n,3}(\theta,P), B_n(\theta,P)$ for all $\theta\in\Theta_0(P)$; and let $\mathcal{P}_n^{MB}$ be the class of distributions P on $(S,\mathcal{S})$ for which (26) and (17) are verified with $M_{n,k}, B_n$ replaced by (respectively) $M_{n,k}(\theta,P), B_n(\theta,P)$ for all $\theta\in\Theta_0(P)$.*[5] *Moreover, suppose that $\sup_{n\geq 1}\beta_n\leq\alpha/3$ and $\log(1/\beta_n)\leq C_1\log n$ whenever moment selection is used. Then there exist positive constants c, C depending only on c_1, C_1 such that*

$$\inf_{P\in\mathcal{P}_n}\inf_{\theta\in\Theta_0(P)}\mathrm{P}_P(\theta\in\mathcal{C}_n(\alpha))\geq 1-\alpha-Cn^{-c}$$

for every pair

$$(\mathcal{P}_n,\mathcal{C}_n)\in\Big\{(\mathcal{P}_n^{SN},\mathcal{C}_n^{SN}),(\mathcal{P}_n^{SN,MS},\mathcal{C}_n^{SN,MS}),(\mathcal{P}_n^{MB},\mathcal{C}_n^{MB}),(\mathcal{P}_n^{MB},\mathcal{C}_n^{MB,MS}),(\mathcal{P}_n^{MB},\mathcal{C}_n^{HB})\Big\}.$$

6. An extension: empirical bootstrap

In this paper, among many bootstrap procedures, we have focused on the multiplier bootstrap (see, for example, Praestgaard and Wellner, 1993, for other forms of bootstrap). Indeed, many other bootstrap procedures are asymptotically equivalent to the multiplier bootstrap, as discussed briefly in Comment 3.2 of Chernozhukov, Chetverikov, and Kato (2013a). For example, we consider here the empirical (or Efron's) bootstrap. Let $X_1^*,\dots,X_n^*$ be i.i.d. draws from the empirical distribution of $X_1^n=\{X_1,\dots,X_n\}$, and construct the bootstrap versions of $\widehat{\mu}_j$ and $\widehat{\sigma}_j^2$ as

$$\widehat{\mu}_j^*=\mathbb{E}_n[X_{ij}^*],\ \widehat{\sigma}_j^{*2}=\mathbb{E}_n[(X_{ij}^*-\widehat{\mu}_j^*)^2].$$

Let

$$W^*=\max_{1\leq j\leq p}\frac{\sqrt{n}(\widehat{\mu}_j^*-\widehat{\mu}_j)}{\widehat{\sigma}_j^*},$$

and consider

$$c^{EB}(\alpha)=\text{conditional }(1-\alpha)\text{-quantile of }W^*\text{ given }X_1^n,$$

[5]For example, $\mathcal{P}_n^{SN}=\{P:\ (26)\text{ is verified, and } M_{n,3}^3(\theta,P)\log^{3/2}(p/\alpha)\leq C_1n^{1/2-c_1},\forall\theta\in\Theta_0(P)\}$.

which we call the EB (Empirical Bootstrap) critical value.

Informally, due to conditional independence of $X_1^*, \dots, X_n^*$, the conditional distribution of W^* should be close to that of W (defined in (15)), so that validity of the EB critical value will follow from Theorem 3.3. Formally we have the following theorem.

Theorem 6.1 (Validity of EB method)**.** *Suppose that there exists a sequence of positive constants D_n such that*

$$\mathrm{P}(|Z_{ij}| \leq D_n,\ 1 \leq \forall i \leq n, 1 \leq \forall j \leq p) = 1. \tag{28}$$

In addition, suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that

$$D_n^2 \log^{7/2}(pn) \leq C_1 n^{1/2 - c_1}. \tag{29}$$

Then there exist positive constants c, C depending only on c_1, C_1 such that under H_0,

$$\mathrm{P}(T > c^{EB}(\alpha)) \leq \alpha + Cn^{-c}.$$

If $\mu_j = 0$ for all $1 \leq j \leq p$, then

$$|\mathrm{P}(T > c^{EB}(\alpha))| \leq \alpha + Cn^{-c}.$$

Moreover, all these bounds hold uniformly with respect to the common distribution of X_i for which (3), (28) and (29) are verified.

Comment 6.1. The conditions in Theorem 6.1 are stronger than those in Theorem 3.3 in that we need to assume that the random variables X_{ij} are bounded, although D_n is allowed to diverge as $n \to \infty$. A detailed analysis of more sophisticated conditions, and the validity of more general exchangeably weighted bootstraps (Praestgaard and Wellner, 1993) will be pursued in the future work.

Appendix A. Proofs

In what follows, let $\phi(\cdot)$ denote the density function of the standard normal distribution, and let $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$ where recall that $\Phi(\cdot)$ is the distribution function of the standard normal distribution.

A.1. **Technical tools.** We state here some technical tools used to prove the theorems. The following lemma states a moderate deviation inequality for self-normalized sums.

Lemma A.1. *Let $\xi_1, \dots, \xi_n$ be independent centered random variables with $\mathrm{E}[\xi_i^2] = 1$ and $\mathrm{E}[|\xi_i|^{2+\nu}] < \infty$ for all $1 \leq i \leq n$ where $0 < \nu \leq 1$. Let $S_n = \sum_{i=1}^n \xi_i, V_n^2 = \sum_{i=1}^n \xi_i^2$, and $D_{n,\nu} = (n^{-1} \sum_{i=1}^n \mathrm{E}[|\xi_i|^{2+\nu}])^{1/(2+\nu)}$. Then uniformly in $0 \leq x \leq n^{\frac{\nu}{2(2+\nu)}}/D_{n,\nu}$,*

$$\left| \frac{\mathrm{P}(S_n/V_n \geq x)}{\bar{\Phi}(x)} - 1 \right| \leq K n^{-\nu/2} D_{n,\nu}^{2+\nu} (1+x)^{2+\nu},$$

where K is a universal constant.

Proof. See Theorem 7.4 in Lai, de la Peña, and Shao (2009) or the original paper Jing, Shao, and Wang (2003). □

The following lemma states a Fuk-Nagaev type inequality, which is a deviation inequality for the maximum of the sum of random vectors from its expectation.

Lemma A.2 (A Fuk-Nagaev type inequality). *Let $X_1, \dots, X_n$ be independent random vectors in $\mathbb{R}^p$. Define $\sigma^2 := \max_{1 \le j \le p} \sum_{i=1}^n \mathrm{E}[X_{ij}^2]$. Then for every $s > 1$ and $t > 0$,*

$$\mathrm{P}\left(\max_{1 \le j \le p} \left| \sum_{i=1}^n (X_{ij} - \mathrm{E}[X_{ij}]) \right| \ge 2\mathrm{E}\left[\max_{1 \le j \le p} \left| \sum_{i=1}^n (X_{ij} - \mathrm{E}[X_{ij}]) \right| \right] + t \right)$$
$$\le e^{-t^2/(3\sigma^2)} + K_s \sum_{i=1}^n \mathrm{E}[\max_{1 \le j \le p} |X_{ij}|^s]/t^s,$$

where K_s is a constant depending only on s.

Proof. See Theorem 3.1 in Einmahl and Li (2008). Note that Einmahl and Li (2008) assumed that $s > 2$ but their proof applies to the case where $s > 1$. More precisely, we apply Theorem 3.1 in Einmahl and Li (2008) with $(B, \|\cdot\|) = (\mathbb{R}^p, |\cdot|_\infty)$ where $|x|_\infty = \max_{1 \le j \le p} |x_j|$ for $x = (x_1, \dots, x_p)^T$, and $\eta = \delta = 1$. The unit ball of the dual of $(\mathbb{R}^p, |\cdot|_\infty)$ is the set of linear functions $\{x = (x_1, \dots, x_p)^T \mapsto \sum_{j=1}^p \lambda_j x_j : \sum_{j=1}^p |\lambda_j| \le 1\}$, and for $\lambda_1, \dots, \lambda_p$ with $\sum_{j=1}^p |\lambda_j| \le 1$, by Jensen's inequality,

$$\sum_{i=1}^n \mathrm{E}[(\sum_{j=1}^p \lambda_j X_{ij})^2] = \sum_{i=1}^n \mathrm{E}[(\sum_{j=1}^p |\lambda_j| \mathrm{sign}(\lambda_j) X_{ij})^2]$$
$$\le \sum_{j=1}^p |\lambda_j| \sum_{i=1}^n \mathrm{E}[X_{ij}^2] \le \max_{1 \le j \le p} \sum_{i=1}^n \mathrm{E}[X_{ij}^2] = \sigma^2,$$

where $\mathrm{sign}(\lambda_j)$ is the sign of λ_j. Hence in this case Λ_n^2 in Theorem 3.1 of Einmahl and Li (2008) is bounded by (and indeed equal to) σ^2. □

In order to use Lemma A.2, we need a suitable bound on the expectation of the maximum. The following lemma is useful for that purpose.

Lemma A.3. *Let $X_1, \dots, X_n$ be independent random vectors in $\mathbb{R}^p$ with $p \ge 2$. Define $M := \max_{1 \le i \le n} \max_{1 \le j \le p} |X_{ij}|$ and $\sigma^2 := \max_{1 \le j \le p} \sum_{i=1}^n \mathrm{E}[X_{ij}^2]$. Then*

$$\mathrm{E}\left[\max_{1 \le j \le p} \left| \sum_{i=1}^n (X_{ij} - \mathrm{E}[X_{ij}]) \right| \right] \le K(\sigma \sqrt{\log p} + \sqrt{\mathrm{E}[M^2]} \log p),$$

where K is a universal constant.

Proof. See Lemma 8 in Chernozhukov, Chetverikov, and Kato (2013b). □

For bounding $\mathrm{E}[M^2]$, we will frequently use the following inequality: let $\xi_1, \dots, \xi_n$ be arbitrary random variables with $\mathrm{E}[|\xi_i|^s] < \infty$ for all $1 \le i \le n$

for some $s \geq 1$. Then

$$\begin{aligned} \mathrm{E}[\max_{1\leq i\leq n} |\xi_i|] &\leq (\mathrm{E}[\max_{1\leq i\leq n} |\xi_i|^s])^{1/s} \\ &\leq (\textstyle\sum_{i=1}^n \mathrm{E}[|\xi_i|^s])^{1/s} \leq n^{1/s} \max_{1\leq i\leq n} (\mathrm{E}[|\xi_i|^s])^{1/s}. \end{aligned}$$

For centered normal random variables $\xi_1, \dots, \xi_n$ with $\sigma^2 = \max_{1\leq i\leq n} \mathrm{E}[\xi_i^2]$, we have

$$\mathrm{E}[\max_{1\leq j\leq p} \xi_i] \leq \sqrt{2\sigma^2 \log p}.$$

See, for example, Proposition 1.1.3 in Talagrand (2003).

Lemma A.4. *Let* $(Y_1, \dots, Y_p)^T$ *be a normal random vector with* $\mathrm{E}[Y_j] = 0$ *and* $\mathrm{E}[Y_j^2] = 1$ *for all* $1 \leq j \leq p$. *(i) For* $\alpha \in (0,1)$, *let* $c_0(\alpha)$ *denote the* $(1-\alpha)$*-quantile of the distribution of* $\max_{1\leq j\leq p} Y_j$. *Then* $c_0(\alpha) \leq \sqrt{2\log p} + \sqrt{2\log(1/\alpha)}$. *(ii) For every* $t \in \mathbb{R}$ *and* $\epsilon > 0$, $\mathrm{P}(|\max_{1\leq j\leq p} Y_j - t| \leq \epsilon) \leq 4\epsilon(\sqrt{2\log p} + 1)$.

Proof. Part (ii) follows from Theorem 3 in Chernozhukov, Chetverikov, and Kato (2013b) together with the fact that

$$\mathrm{E}[\max_{1\leq j\leq p} Y_j] \leq \sqrt{2\log p}. \tag{30}$$

For part (i), by the Borell-Sudakov-Tsirelson inequality (see Theorem A.2.1 in van der Vaart and Wellner (1996)), for every $r > 0$,

$$\mathrm{P}(\max_{1\leq j\leq p} Y_j \geq \mathrm{E}[\max_{1\leq j\leq p} Y_j] + r) \leq e^{-r^2/2},$$

by which we have

$$c_0(\alpha) \leq \mathrm{E}[\max_{1\leq j\leq p} Y_j] + \sqrt{2\log(1/\alpha)}. \tag{31}$$

Combining (31) and (30) leads to the desired result. □

A.2. **Proof of Theorem 3.1.** The first assertion follows from inequality (8) and Lemma A.1 with $\nu = 1$. To prove the second assertion, we first note the well known fact that $1 - \Phi(t) \leq e^{-t^2/2}$ for $t > 0$, by which we have $\Phi^{-1}(1 - \alpha/p) \leq \sqrt{2\log(p/\alpha)}$.[6] Hence if $M_{n,3}^3 \log^{3/2}(p/\alpha) \leq C_1 n^{1/2-c_1}$, it is straightforward to verify that the right side on (10) is bounded by Cn^{-c_1} for some constant C depending only on c_1, C_1. □

[6]The inequality $1-\Phi(t) \leq e^{-t^2/2}$ for $t > 0$ can be proved by using Markov's inequality, $\mathrm{P}(\xi > t) \leq e^{-\lambda t}\mathrm{E}[e^{\lambda\xi}]$ for $\lambda > 0$ with $\xi \sim N(0,1)$, and optimizing the bound with respect to $\lambda > 0$; there is a sharper inequality, namely $1 - \Phi(t) \leq e^{-t^2/2}/2$ for $t > 0$ (see, for example, Proposition 2.1 in Dudley, 1999), but we do not need this sharp inequality in this paper.

A.3. **Proof of Theorem 3.2.** We first prove the following technical lemma. Recall that $B_n = (\mathrm{E}[\max_{1\le j\le p} Z_{1j}^4])^{1/4}$.

Lemma A.5. *For every* $0 < c < 1$,

$$\mathrm{P}\left(\max_{1\le j\le p} |(\widehat{\sigma}_j/\sigma_j) - 1| > K(n^{-(1-c)/2}B_n^2 \log p + n^{-3/2}B_n^2 \log^2 p)\right) \le K' n^{-c},$$

where K, K' *are universal constants.*

Proof. Here $K_1, K_2, \dots$ denote universal positive constants. Note that for $a > 0$,

$$|\sqrt{a} - 1| = \frac{|a-1|}{\sqrt{a}+1} \le |a-1|,$$

so that for $r > 0$,

$$\mathrm{P}\left(\max_{1\le j\le p} |(\widehat{\sigma}_j/\sigma_j) - 1| > r\right) \le \mathrm{P}\left(\max_{1\le j\le p} |(\widehat{\sigma}_j^2/\sigma_j^2) - 1| > r\right).$$

Using the expression

$$\widehat{\sigma}_j^2/\sigma_j^2 - 1 = (\mathbb{E}_n[Z_{ij}^2] - 1) - (\mathbb{E}_n[Z_{ij}])^2,$$

we have

$$\begin{aligned}&\mathrm{P}\left(\max_{1\le j\le p} |\widehat{\sigma}_j^2/\sigma_j^2 - 1| > r\right)\\ &\le \mathrm{P}\left(\max_{1\le j\le p} |\mathbb{E}_n[Z_{ij}^2] - 1| > r/2\right) + \mathrm{P}\left(\max_{1\le j\le p} |\mathbb{E}_n[Z_{ij}]| > \sqrt{r/2}\right).\end{aligned}$$

We wish to bound the two terms on the right side by using the Fuk-Nagaev inequality (Lemma A.2) combined with the maximal inequality of Lemma A.3.

By Lemma A.3 (with the crude bounds $\mathrm{E}[Z_{1j}^4] \le B_n^4$ and $\mathrm{E}[\max_{i,j} Z_{ij}^4] \le nB_n^4$), we have

$$\mathrm{E}\left[\max_{1\le j\le p} |\mathbb{E}_n[Z_{ij}^2] - 1|\right] \le K_1 B_n^2 (\log p)/\sqrt{n},$$

so that by Lemma A.2, for every $t > 0$,

$$\mathrm{P}\left(\max_{1\le j\le p} |\mathbb{E}_n[Z_{ij}^2] - 1| > 2K_1 B_n^2 (\log p)/\sqrt{n} + t\right) \le e^{-nt^2/(3B_n^4)} + K_2 t^{-2} n^{-1} B_n^4.$$

Taking $t = n^{-(1-c)/2} B_n^2$ with $0 < c < 1$, the right side becomes $e^{-n^c/3} + K_2 n^{-c} \le K_3 n^{-c}$. Hence we have

$$\mathrm{P}\left(\max_{1\le j\le p} |\mathbb{E}_n[Z_{ij}^2] - 1| > K_4 n^{-(1-c)/2} B_n^2 (\log p)\right) \le K_3 n^{-c}. \tag{32}$$

Similarly, using Lemma A.3, we have

$$\mathrm{E}\left[\max_{1\le j\le p} |\mathbb{E}_n[Z_{ij}]|\right] \le K_5 (n^{-1/2}\sqrt{\log p} + n^{-3/4} B_n \log p),$$

so that by Lemma A.2, for every $t > 0$,

$$\mathrm{P}\left(\max_{1\le j\le p}|\mathbb{E}_n[Z_{ij}]| > 2K_5(n^{-1/2}\sqrt{\log p} + n^{-3/4}B_n\log p) + t\right)$$
$$\le e^{-nt^2/3} + K_6 t^{-4} n^{-3} B_n^4.$$

Taking $t = n^{-1/4}B_n$, the right side becomes $e^{-n^{1/2}B_n/3} + K_6 n^{-2} \le K_7 n^{-2}$. Hence we have

$$\mathrm{P}\left(\max_{1\le j\le p}|\mathbb{E}_n[Z_{ij}]| > K_8(n^{-1/4}B_n\sqrt{\log p} + n^{-3/4}B_n\log p)\right) \le K_7 n^{-2}. \quad (33)$$

Combining (32) and (33) leads to the desired result. □

Proof of Theorem 3.2. Here c, C denote generic positive constants depending only on c_1, C_1; their values may change from place to place. Define

$$J_1 = \{j \in \{1,\dots,p\} : \sqrt{n}\mu_j > -\sigma_j c^{SN}(\beta_n)\},\ J_1^c = \{1,\dots,p\}\backslash J_1. \quad (34)$$

For $k \ge 1$, let

$$c^{SN,MS}(\alpha,k) = \frac{\Phi^{-1}(1-(\alpha-2\beta_n)/k)}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta_n)/k)^2/n}}.$$

Note that $c^{SN,MS}(\alpha) = c^{SN,MS}(\alpha,\widehat{k})$ when $\widehat{k}\ge 1$. We divide the proof into several steps.

Step 1. We wish to prove that with probability larger than $1-\beta_n-Cn^{-c}$, $\widehat{\mu}_j \le 0$ for all $j \in J_1^c$.

Observe that

$$\widehat{\mu}_j > 0, \exists j \in J_1^c \Rightarrow \sqrt{n}(\widehat{\mu}_j - \mu_j) > \sigma_j c^{SN}(\beta_n), 1 \le \exists j \le p,$$

so that it is enough to prove that

$$\mathrm{P}\left(\max_{1\le j\le p}\left[\sqrt{n}(\widehat{\mu}_j-\mu_j) - \sigma_j c^{SN}(\beta_n)\right] > 0\right) \le \beta_n + Cn^{-c}. \quad (35)$$

Since whenever $\sigma_j/\widehat{\sigma}_j - 1 \ge -r$ for some $0<r<1$,

$$\sigma_j = \widehat{\sigma}_j(1+(\sigma_j/\widehat{\sigma}_j-1)) \ge \widehat{\sigma}_j(1-r),$$

the left side on (35) is bounded by

$$\mathrm{P}\left(\max_{1\le j\le p}\sqrt{n}(\widehat{\mu}_j-\mu_j)/\widehat{\sigma}_j > (1-r)c^{SN}(\beta_n)\right)$$
$$+\,\mathrm{P}\left(\max_{1\le j\le p}|(\sigma_j/\widehat{\sigma}_j)-1| > r\right), \quad (36)$$

where $0<r<1$ is arbitrary.

Take $r = r_n = n^{-(1-c_1)/2}B_n^2\log p$. Then $r_n < 1$ for large n, and since

$$|a-1| \le \frac{r}{r+1} \Rightarrow |a^{-1}-1| \le r,$$

we see that by Lemma A.5, the second term in (36) is bounded by Cn^{-c}.

Consider the first term in (36). It is not difficult to see that

$$
\begin{aligned}
&\mathrm{P}\left(\max_{1\le j\le p}\sqrt{n}(\widehat{\mu}_j-\mu_j)/\widehat{\sigma}_j>(1-r)c^{SN}(\beta_n)\right)\\
&\le \mathrm{P}\left(\max_{1\le j\le p}U_j>(1-r)\Phi^{-1}(1-\beta_n/p)\right)\\
&\le \sum_{j=1}^p \mathrm{P}\left(U_j>(1-r)\Phi^{-1}(1-\beta_n/p)\right).
\end{aligned}
$$

Note that $(1-r)\Phi^{-1}(1-\beta_n/p)\le\sqrt{2\log(p/\beta_n)}\le n^{1/6}/M_{n,3}$ for large n. Hence, by Lemma A.1, the far right side is bounded by

$$
\begin{aligned}
&p\bar{\Phi}\left((1-r)\Phi^{-1}(1-\beta_n/p)\right)\left[1+n^{-1/2}CM_{n,3}^3\left\{1+(1-r)\Phi^{-1}(1-\beta_n/p)\right\}^3\right]\\
&\le p\bar{\Phi}\left((1-r)\Phi^{-1}(1-\beta_n/p)\right)\left[1+n^{-1/2}CM_{n,3}^3\{1+\Phi^{-1}(1-\beta_n/p)\}^3\right].
\end{aligned}
$$

Observe that $n^{-1/2}M_{n,3}^3\{1+\Phi^{-1}(1-\beta_n/p)\}^3\le Cn^{-c_1}$. Moreover, putting $\xi=\Phi^{-1}(1-\beta_n/p)$, we have

$$
\begin{aligned}
p\bar{\Phi}\left((1-r)\xi\right)&=\beta_n+rp\xi\phi\left((1-r')\xi\right)\quad(\exists r'\in[0,r])\\
&\le \beta_n+rp\xi\phi\left((1-r)\xi\right).
\end{aligned}
$$

Using the inequality $(1-r)^2\xi^2=\xi^2+r^2\xi^2-2r\xi^2\ge\xi^2-2r\xi^2$, we have $\phi\left((1-r)\xi\right)\le e^{rx^2}\phi(\xi)$. Since $\beta_n<\alpha/2<1/4$ and $p\ge 2$, we have $\xi\ge\Phi^{-1}(1-1/8)>1$, so that by Proposition 2.1 in Dudley (1999), we have $\phi(\xi)\le 2\xi(1-\Phi(\xi))=2\xi\beta_n/p$.[7] Hence

$$
p\bar{\Phi}\left((1-r)\xi\right)\le\beta_n(1+2r\xi^2e^{r\xi^2}).
$$

Recall that we have taken $r=r_n=n^{-(1-c_1)/2}B_n^2\log p$, so that

$$
r\xi^2\le 2n^{-(1-c_1)/2}B_n^2\log^2(p/\beta_n)\le Cn^{-c_1/2}.
$$

Therefore, the first term in (36) is bounded by β_n+Cn^{-c} for large n. The conclusion of Step 1 is verified for large n and hence for all n by adjusting the constant C.

Step 2. We wish to prove that with probability larger than $1-\beta_n-Cn^{-c}$, $\widehat{J}_{SN}\supset J_1$.

Observe that

$$
\mathrm{P}(\widehat{J}_{SN}\not\supset J_1)\le\mathrm{P}\left(\max_{1\le j\le p}\left[\sqrt{n}(\mu_j-\widehat{\mu}_j)-(2\widehat{\sigma}_j-\sigma_j)c^{SN}(\beta_n)\right]>0\right). \quad (37)
$$

Since whenever $1-\sigma_j/\widehat{\sigma}_j\ge -r$ for some $0<r<1$,

$$
2\widehat{\sigma}_j-\sigma_j=\widehat{\sigma}_j(1+(1-\sigma_j/\widehat{\sigma}_j))\ge\widehat{\sigma}_j(1-r),
$$

[7]Note that the second part of Proposition 2.1 in Dudley (1999) asserts that $\phi(t)/t\le\mathrm{P}(|N(0,1)|>t)=2(1-\Phi(t))$ when $t\ge 1$, so that $\phi(t)\le 2t(1-\Phi(t))$.

the right side on (37) is bounded by

$$\mathrm{P}\left(\max_{1\le j\le p}\sqrt{n}(\mu_j-\widehat{\mu}_j)/\widehat{\sigma}_j>(1-r)c^{SN}(\beta_n)\right)$$
$$+\mathrm{P}\left(\max_{1\le j\le p}|(\sigma_j/\widehat{\sigma}_j)-1|>r\right),$$

where $0<r<1$ is arbitrary. By the proof of Step 1, we see that the sum of these terms is bounded by β_n+Cn^{-c} with suitable r, which leads to the conclusion of Step 2.

Step 3. We are now in position to finish the proof of Theorem 3.2. Consider first the case where $J_1=\emptyset$. Then by Step 1, with probability larger than β_n+Cn^{-c}, $T\le 0$, so that $\mathrm{P}(T>c^{SN,MS}(\alpha))\le\beta_n+Cn^{-c}\le\alpha+Cn^{-c}$. Suppose now that $|J_1|\ge 1$. Observe that

$$\{T>c^{SN,MS}(\alpha)\}\cap\{\widehat{\mu}_j\le 0,\forall j\in J_1^c\}\subset\{\max_{j\in J_1}(\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j)>c^{SN,MS}(\alpha)\}.$$

Moreover, as $c^{SN,MS}(\alpha,k)$ is non-decreasing in k,

$$\{\max_{j\in J_1}(\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j)>c^{SN,MS}(\alpha)\}\cap\{\widehat{J}_{SN}\supset J_1\}$$
$$\subset\{\max_{j\in J_1}(\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j)>c^{SN,MS}(\alpha,|J_1|)\}.$$

Therefore, by Steps 1 and 2, we have

$$\begin{aligned}&\mathrm{P}(T>c^{SN,MS}(\alpha))\\&\le\mathrm{P}(\max_{j\in J_1}(\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j)>c^{SN,MS}(\alpha,|J_1|))+2\beta_n+Cn^{-c}\\&\le\mathrm{P}(\max_{j\in J_1}\sqrt{n}(\widehat{\mu}_j-\mu_j)/\widehat{\sigma}_j>c^{SN,MS}(\alpha,|J_1|))+2\beta_n+Cn^{-c}.\end{aligned}\tag{38}$$

By Theorem 3.1, we see that

$$\mathrm{P}(\max_{j\in J_1}\sqrt{n}(\widehat{\mu}_j-\mu_j)/\widehat{\sigma}_j>c^{SN,MS}(\alpha,|J_1|))\le\alpha-2\beta_n+Cn^{-c}.\tag{39}$$

Combining (38) and (39) completes the proof of the theorem. □

A.4. **Proof of Theorem 3.3.** Here c,C denote generic positive constants depending only on c_1,C_1; their values may change from place to place. Define

$$\bar{T}=\max_{1\le j\le p}\frac{\sqrt{n}(\widehat{\mu}_j-\mu_j)}{\widehat{\sigma}_j},\ \text{and}\ T_0=\max_{1\le j\le p}\frac{\sqrt{n}(\widehat{\mu}_j-\mu_j)}{\sigma_j}.$$

Moreover, define

$$\bar{W}=\max_{1\le j\le p}\frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_j)]}{\sigma_j},\quad\text{and}\quad W_0=\max_{1\le j\le p}\frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\mu_j)]}{\sigma_j}.$$

We begin with noting that under H_0, $\mathrm{P}(T>c^{MB}(\alpha))\le\mathrm{P}(\bar{T}>c^{MB}(\alpha))$, and if all the moment inequalities are binding, that is, $\mu_j=0$ for all $1\le j\le p$, then $\mathrm{P}(T>c^{MB}(\alpha))=\mathrm{P}(\bar{T}>c^{MB}(\alpha))$. Hence, by Corollary 3.1 of Chernozhukov, Chetverikov, and Kato (2013a) (applied with $x_i=Z_i$), the conclusion of the theorem follows if we can prove that there exist sequences $\{\zeta_{n1}\}$ and $\{\zeta_{n2}\}$ of positive constants with $\zeta_{n1}\sqrt{\log p}+\zeta_{n2} \leq Cn^{-c}$ such that

$$\mathrm{P}(|\bar{T}-T_0| > \zeta_{n1}) \leq Cn^{-c},\ \mathrm{P}(\mathrm{P}(|W-W_0| > \zeta_{n1} \mid X_1^n) > \zeta_{n2}) \leq Cn^{-c}. \quad (40)$$

We wish to verify these conditions with

$$\zeta_{n1} = n^{-(1-c_1)/2} B_n^2 \log^{3/2} p, \text{ and } \zeta_{n2} = C' n^{-c'},$$

where c', C' are suitable positive constants that depend only on c_1, C_1. We note that because of the assumption that $B_n^4 \log^7(pn) \leq C_1 n^{1/2-c_1}$, these choices satisfy $\zeta_{n1}\sqrt{\log p} + \zeta_{n2} \leq Cn^{-c}$.

We first verify the first part of (40). Observe that

$$|\bar{T} - T_0| \leq \max_{1\leq j\leq p} |(\sigma_j/\widehat{\sigma}_j) - 1| \times \max_{1\leq j\leq p} |\sqrt{n}\mathbb{E}_n[Z_{ij}]|.$$

By Lemma A.5 and the simple fact that $|a-1| \leq r/(r+1) \Rightarrow |a^{-1}-1| \leq r$ $(r > 0)$, we have

$$\mathrm{P}\left(\max_{1\leq j\leq p} |(\sigma_j/\widehat{\sigma}_j) - 1| > n^{-1/2+c_1/4} B_n^2 \log p\right) \leq Cn^{-c}.$$

Moreover,

$$\mathrm{P}(\max_{1\leq j\leq p} |\sqrt{n}\mathbb{E}_n[Z_{ij}]| > n^{c_1/4}\sqrt{\log p}) \leq Cn^{-c}.$$

Hence the first part of (40) is verified (note that $n^{-1/2+c_1/4}B_n^2(\log p) \times n^{c_1/4}\sqrt{\log p} = \zeta_{n1}$).

To verify the second part of (40), define A_n by the event such that

$$A_n = \left\{\max_{1\leq j\leq p} |(\widehat{\sigma}_j/\sigma_j) - 1| \leq n^{-1/2+c_1/4} B_n^2 \log p\right\} \bigcap \left\{\max_{1\leq j\leq p} |\sqrt{n}\mathbb{E}_n[Z_{ij}]| \leq n^{c_1/4}\sqrt{\log p}\right\}.$$

By the previous step, we see that $\mathrm{P}(A_n) > 1 - Cn^{-c}$. Observe that

$$\mathrm{P}(|W-W_0| > \zeta_{n1} \mid X_1^n) \leq \mathrm{P}(|W-\bar{W}| > \zeta_{n1}/2 \mid X_1^n) + \mathrm{P}(|\bar{W}-W_0| > \zeta_{n1}/2 \mid X_1^n).$$

Consider the first term on the right side. Observe that

$$|W - \bar{W}| \leq \max_{1\leq j\leq p} |(\widehat{\sigma}_j/\sigma_j) - 1| \times |W|.$$

Conditional on the data X_1^n, the vector $(\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_j)/\widehat{\sigma}_j])_{1\leq j\leq p}$ is normal with mean zero and all the diagonal elements of the covariance matrix are one. Hence

$$\mathrm{E}[|W| \mid X_1^n] \leq \sqrt{2\log(2p)},$$

so that by Markov's inequality, on the event A_n,

$$\mathrm{P}(|W - \bar{W}| > \zeta_{n1}/2 \mid X_1^n) \leq (2/\zeta_{n1}) \max_{1\leq j\leq p} |(\widehat{\sigma}_j/\sigma_j) - 1| \times \mathrm{E}[|W| \mid X_1^n],$$

which is bounded by $Cn^{-c_1/4}$. On the other hand, observe that

$$|\bar{W} - W_0| \leq |\mathbb{E}_n[\epsilon_i]| \times \max_{1\leq j\leq p} |\sqrt{n}\mathbb{E}_n[Z_{ij}]|.$$

Hence by Markov's inequality, on the event A_n,

$$\mathrm{P}(|\bar{W} - W_0| > \zeta_{n1}/2) \le C\zeta_{n1}^{-1} n^{-1/2} \times n^{c_1/4}\sqrt{\log p} \le Cn^{-c_1/4}.$$

Therefore, the second part of (40) is verified. □

A.5. **Proof of Theorem 3.4.** We first prove the following technical lemma. Define

$$\widehat{\Delta} = \max_{1\le j,k\le p}\left|\mathbb{E}_n\left[\frac{(X_{ij}-\widehat{\mu}_j)}{\widehat{\sigma}_j}\frac{(X_{ik}-\widehat{\mu}_k)}{\widehat{\sigma}_k}\right] - \mathrm{E}[Z_{1j}Z_{1k}]\right|.$$

Recall that $B_n = (\mathrm{E}[\max_{1\le j\le p} Z_{1j}^4])^{1/4}$.

Lemma A.6. *Suppose that there exists a constant C_1 such that $\log p \le C_1 n$. Then for every $0 < c < 1$, there exist positive constants C, C' depending only on c, C_1 such that*

$$\mathrm{P}\left(\widehat{\Delta} > C' n^{-(1-c)/2} B_n^2 \log p\right) \le Cn^{-c},$$

whenever $C' n^{-(1-c)/2} B_n^2 \log p \le 1/2$.

Proof. Here C, C' denote generic positive constants depending only on c, C_1; their values may change from place to place. Observe that

$$\begin{aligned}
&\mathbb{E}_n\left[\frac{(X_{ij}-\widehat{\mu}_j)}{\widehat{\sigma}_j}\frac{(X_{ik}-\widehat{\mu}_k)}{\widehat{\sigma}_k}\right] - \mathrm{E}[Z_{1j}Z_{1k}] \\
&= \frac{1}{(\widehat{\sigma}_j/\sigma_j)(\widehat{\sigma}_k/\sigma_k)}(\mathbb{E}_n[Z_{ij}Z_{ik}] - \mathrm{E}[Z_{1j}Z_{1k}]) - \frac{1}{(\widehat{\sigma}_j/\sigma_j)(\widehat{\sigma}_k/\sigma_k)}\mathbb{E}_n[Z_{ij}]\mathbb{E}_n[Z_{ik}] \\
&\quad + \left(\frac{1}{(\widehat{\sigma}_j/\sigma_j)(\widehat{\sigma}_k/\sigma_k)} - 1\right)\mathrm{E}[Z_{1j}Z_{ik}].
\end{aligned}$$

Hence, since $|\mathrm{E}[Z_{1j}Z_{1k}]| \le 1$, we have

$$\begin{aligned}
\widehat{\Delta} &\le \max_{1\le j,k\le p}\left|\frac{1}{(\widehat{\sigma}_j/\sigma_j)(\widehat{\sigma}_k/\sigma_k)}(\mathbb{E}_n[Z_{ij}Z_{ik}] - \mathrm{E}[Z_{1j}Z_{1k}])\right| \\
&\quad + \max_{1\le j,k\le p}\left|\frac{1}{(\widehat{\sigma}_j/\sigma_j)(\widehat{\sigma}_k/\sigma_k)}\mathbb{E}_n[Z_{ij}]\mathbb{E}_n[Z_{ik}]\right| + \max_{1\le j,k\le p}\left|\frac{1}{(\widehat{\sigma}_j/\sigma_j)(\widehat{\sigma}_k/\sigma_k)} - 1\right| \\
&=: I + II + III.
\end{aligned}$$

By Lemma A.5, we see that with probability larger than $1 - Cn^{-c}$,

$$\max_{1\le j\le p}|(\widehat{\sigma}_j/\sigma_j) - 1| \le C' n^{-(1-c)/2} B_n^2 \log p,$$

Hence as long as $C' n^{-(1-c)/2} B_n^2 \log p \le 1/2$, with probability larger than $1 - Cn^{-c}$,

$$\widehat{\sigma}_j/\sigma_j \ge 1/2, \ 1 \le \forall j \le p, \text{ and } III \le C' n^{-(1-c)/2} B_n^2 \log p.$$

On the event $\widehat{\sigma}_j/\sigma_j \ge 1/2, \ 1 \le \forall j \le p$, we have

$$I \le 4\max_{1\le j,k\le p}|\mathbb{E}_n[Z_{ij}Z_{ik}] - \mathrm{E}[Z_{1j}Z_{1k}]|, \ II \le 4\max_{1\le j\le p}(\mathbb{E}_n[Z_{ij}])^2.$$

By inequality (33) in the proof of Lemma A.5, we see that with probability larger than $1 - Cn^{-2}$, $\max_{1\le j\le p}(\mathbb{E}_n[Z_{ij}])^2 \le C'n^{-1/2}B_n^2 \log p$. Therefore, the conclusion of the lemma will follow if we can prove that with probability larger than $1 - Cn^{-c}$,

$$\widehat{\Delta}' := \max_{1\le j,k\le p} |\mathbb{E}_n[Z_{ij}Z_{ik}] - \mathrm{E}[Z_{1j}Z_{1k}]| \le C'n^{-(1-c)/2}B_n^2 \log p.$$

By Lemma A.2, for every $t > 0$,

$$\mathrm{P}(\widehat{\Delta}' > 2\mathrm{E}[\widehat{\Delta}'] + t) \le e^{-nt^2/(3B_n^4)} + Kt^{-2}n^{-1}B_n^4.$$

Moreover, by Lemma A.3,

$$\mathrm{E}[\widehat{\Delta}'] \le K'B_n^2(\log n)/\sqrt{n}.$$

Hence by taking $t = n^{-(1-c)/2}B_n^2$, we see that

$$\mathrm{P}(\widehat{\Delta}' > C'n^{-(1-c)/2}B_n^2 \log p) \le Cn^{-c}.$$

This completes the proof. □

Proof of Theorem 3.4. Here c, C denote generic positive constants depending only on c_1, C_1; their values may change from place to place. Let

$$(Y_1, \dots, Y_p)^T \sim N(0, \mathrm{E}[Z_1 Z_1^T]).$$

For $\gamma \in (0, 1)$, denote by $c_0(\gamma)$ the $(1 - \gamma)$-quantile of the distribution of $\max_{1\le j\le p} Y_j$. We divide the proof into several steps.

Step 1. We wish to prove that with probability larger than $1 - Cn^{-c}$, $c^{MB}(\beta_n) \ge c_0(\beta_n+\nu_n)$, where $\nu_n = C'n^{-c'}$ and where c', C' are some positive constant depending only on c_1, C_1.

Recall that conditional on X_1^n, the vector $(\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \widehat{\mu}_j)/\widehat{\sigma}_j])_{1\le j\le p}$ is normal with mean zero and covariance matrix

$$\mathbb{E}_n\left[\frac{(X_{ij} - \widehat{\mu}_j)}{\widehat{\sigma}_j}\frac{(X_{ik} - \widehat{\mu}_k)}{\widehat{\sigma}_k}\right],\ j, k = 1, \dots, p.$$

Hence by Theorem 2 in Chernozhukov, Chetverikov, and Kato (2013b),

$$\sup_{t\in\mathbb{R}} |\mathrm{P}(W \le t \mid X_1^n) - \mathrm{P}(\max_{1\le j\le p} Y_j \le t)| \le K\widehat{\Delta}^{1/3}(1 \vee \log(p/\widehat{\Delta}))^{2/3},$$

where K is a universal constant. Because of the assumption that $B_n^2 \log^{7/2}(pn) \le C_1 n^{1/2-c_1}$ and Lemma A.6, we see that with probability larger than $1-Cn^{-c}$, the right side is bounded by $C'n^{-c'}(=: \nu_n)$ with suitable constants c', C' that depend only on c_1, C_1. This implies that with probability larger than $1 - Cn^{-c}$,

$$\mathrm{P}(\max_{1\le j\le p} Y_j \le t) \ge \mathrm{P}(W \le t \mid X_1^n) - \nu_n,\ \forall t \in \mathbb{R}.$$

Since the conditional distribution of W has no point masses, we have

$$\mathrm{P}(\max_{1\le j\le p} Y_j \le t)|_{t=c^{MB}(\beta_n)} \ge 1 - \beta_n - \nu_n,$$

which implies that $c^{MB}(\beta_n) \ge c_0(\beta_n + \nu_n)$.

Step 2. Define

$$J_2 = \{j \in \{1,\dots,p\} : \sqrt{n}\mu_j > -\sigma_j c_0(\beta_n + \nu_n)\},\ J_2^c = \{1,\dots,p\}\backslash J_2.$$

We wish to prove that with probability larger than $1-\beta_n - Cn^{-c}$, $\widehat{\mu}_j \leq 0$ for all $j \in J^c$.

Like in the proof of Theorem 3.2, observe that

$$\widehat{\mu}_j > 0, \exists j \in J_2^c \Rightarrow \sqrt{n}(\widehat{\mu}_j - \mu_j) > \sigma_j c_0(\beta_n + \nu_n), 1 \leq \exists j \leq p,$$

so that it is enough to prove that

$$\mathrm{P}\left(\max_{1\leq j\leq p} \frac{\sqrt{n}(\widehat{\mu}_j - \mu_j)}{\sigma_j} > c_0(\beta_n + \nu_n) \right) \leq \beta_n + Cn^{-c}.$$

But this follows from Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2013a) (and the fact that $\nu_n = C'n^{-c'}$). This concludes Step 2.

Step 3. We wish to prove that with probability larger than $1-\beta_n-Cn^{-c}$, $\widehat{J}_{MB} \supset J_2$.

Like in the proof of Theorem 3.2, observe that

$$\begin{aligned}&\mathrm{P}(\widehat{J}_{MB} \not\supset J_2)\\ &\quad\leq \mathrm{P}\left(\max_{1\leq j\leq p} \left[\sqrt{n}(\mu_j - \widehat{\mu}_j) - (2\widehat{\sigma}_j c^{MB}(\beta_n) - \sigma_j c_0(\beta_n + \nu_n)) \right] > 0 \right).\end{aligned}$$

Since whenever $c^{MB}(\beta_n) \geq c_0(\beta_n+\nu_n)$ and $\widehat{\sigma}_j/\sigma_j - 1 \geq -r/2$ for some $r > 0$,

$$\begin{aligned}2\widehat{\sigma}_j c^{MB}(\beta_n) - \sigma_j c_0(\beta_n+\nu_n) &\geq (2\widehat{\sigma}_j - \sigma_j)c_0(\beta_n+\nu_n)\\ &= \sigma_j(1+2(\widehat{\sigma}_j/\sigma_j - 1))c_0(\beta_n+\nu_n) \geq (1-r)\sigma_j c_0(\beta_n+\nu_n),\end{aligned}$$

we have

$$\begin{aligned}\mathrm{P}(\widehat{J}_{MB} \not\supset J_2) &\leq \mathrm{P}\left(\max_{1\leq j\leq p} \frac{\sqrt{n}(\mu_j - \widehat{\mu}_j)}{\sigma_j} > (1-r)c_0(\beta_n+\nu_n) \right)\\ &\quad + \mathrm{P}\left(c^{MB}(\beta_n) < c_0(\beta_n+\nu_n)\right) + \mathrm{P}\left(\max_{1\leq j\leq p} |(\widehat{\sigma}_j/\sigma_j) - 1| > r/2 \right).\end{aligned}$$

By Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2013a), the first term on the right side is bounded by

$$\mathrm{P}(\max_{1\leq j\leq p} Y_j > (1-r)c_0(\beta_n+\nu_n)) + Cn^{-c}.$$

Moreover, by Lemma A.4,

$$\begin{aligned}&\mathrm{P}(\max_{1\leq j\leq p} Y_j > (1-r)c_0(\beta_n+\nu_n))\\ &\quad\leq \beta_n + \nu_n + 4r\left(\sqrt{2\log p} + 1)(\sqrt{2\log p} + \sqrt{2\log(1/(\beta_n+\nu_n))}\right),\end{aligned}$$

which is bounded by $\beta_n + \nu_n + Cr\log(pn)$. Thus, using Step 1, we see that

$$\mathrm{P}(\widehat{J}_{MB} \not\supset J_2) \leq \beta_n + \mathrm{P}\left(\max_{1\leq j\leq p} |(\widehat{\sigma}_j/\sigma_j) - 1| > r/2 \right) + C(r\log(pn) + n^{-c}).$$

Choosing $r = r_n = n^{-(1-c_1)/2} B_n^2 \log p$, we see that, by Lemma A.5, the second term on the right side is bounded by Cn^{-c}, and

$$r \log(pn) \leq n^{-(1-c_1)/2} B_n^2 \log^2(pn) \leq C_1 n^{-c_1/2},$$

because of the assumption that $B_n^2 \log^{7/2}(pn) \leq C_1 n^{1/2-c_1}$. This leads to the conclusion of Step 3.

Step 4. We are now in position to finish the proof of the theorem. Note that when $J_2 = \emptyset$, by Step 2 we have that $T \leq 0$ with probability larger than $1 - \beta_n - Cn^{-c}$. But as $c^{MB,MS}(\alpha) \geq 0$, we have $\mathrm{P}(T > c^{MB,MS}(\alpha)) \leq \beta_n + Cn^{-c} \leq \alpha + Cn^{-c}$. Consider the case where $J_2 \neq \emptyset$. Define $c^{MB,MS}(\alpha, J_2)$ by the same bootstrap procedure as $c^{MB,MS}(\alpha)$ with $\widehat{J}_{MB}$ replaced by J_2. Note that $c^{MB,MS}(\alpha) \geq c^{MB,MS}(\alpha, J_2)$ on the event $\widehat{J}_{MB} \supset J_2$. Therefore, arguing as in Step 3 of the proof of Theorem 3.2,

$$\begin{aligned}
\mathrm{P}(T > c^{MB,MS}(\alpha)) &\leq \mathrm{P}(\max_{j \in J_2} \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j > c^{MB,MS}(\alpha)) + \beta_n + Cn^{-c} \\
&\leq \mathrm{P}(\max_{j \in J_2} \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j > c^{MB,MS}(\alpha, J_2)) + 2\beta_n + Cn^{-c} \\
&\leq \mathrm{P}(\max_{j \in J_2} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j > c^{MB,MS}(\alpha, J_2)) + 2\beta_n + Cn^{-c} \\
&\leq \alpha - 2\beta_n + 2\beta_n + Cn^{-c} = \alpha + Cn^{-c}.
\end{aligned}$$

Moreover, when $\mu_j = 0$ for all $1 \leq j \leq p$, we have $J_2 = \{1, \dots, p\}$. Hence by Step 3, $c^{MB,MS}(\alpha) = c^{MB,MS}(\alpha, J_2)$ with probability larger than $1 - \beta_n - Cn^{-c}$. Therefore,

$$\begin{aligned}
\mathrm{P}(T > c^{MB,MS}(\alpha)) &= \mathrm{P}(\max_{1 \leq j \leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j > c^{MB,MS}(\alpha)) \\
&\geq \mathrm{P}(\max_{1 \leq j \leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j > c^{MB,MS}(\alpha, J_2)) - \beta_n - Cn^{-c} \\
&\geq \alpha - 3\beta_n - Cn^{-c}.
\end{aligned}$$

The last assertion follows trivially. This completes the proof of the theorem. □

A.6. **Proof of Theorem 3.5.** Recall the set $J_1 \subset \{1, \dots, p\}$ defined in (34). By Steps 1 and 2 in the proof of Theorem 3.2, we see that

$$\mathrm{P}(\widehat{\mu}_j \leq 0, \forall j \in J_1^c) > 1 - \beta_n - Cn^{-c}, \ \mathrm{P}(\widehat{J}_{SN} \supset J_1) > 1 - \beta_n - Cn^{-c},$$

where c, C are some positive constants depending only on c_1, C_1. The rest of the proof is completely analogous to Step 4 in the proof of Theorem 3.4 and hence omitted. □

A.7. **Proof of Lemma 4.1.** We first assume that $\sigma_1^2, \dots, \sigma_p^2$ are known. Since $\bar{X} := \mathbb{E}_n[X_i]$ is a sufficient statistic for μ, we only have to consider a test that depends only on $\bar{X}$. Let $\phi_n : \mathbb{R}^p \to [0,1], \bar{X} \mapsto \phi_n(\bar{X})$, be a test such that $\mathrm{E}_\mu[\phi_n(\bar{X})] \leq \alpha$ for all $\mu \in \mathbb{R}^p$ with $\max_{1 \leq j \leq p} \mu_j \leq 0$. Let $\mu[j]$ be the vector in $\mathbb{R}^p$ such that only the j-th element is nonzero and $\theta\sigma_j$. Denote by $\mathrm{E}_0[\cdot]$ the expectation under $\mu = 0$, and denote by $\mathrm{E}_j[\cdot]$ the expectation under $\mu = \mu[j]$. Then we have

$$\inf_{\max_{1\le j\le p}(\mu_j/\sigma_j)\ge\theta} \mathrm{E}_\mu[\phi_n(\bar{X})] - \alpha \le \frac{1}{p}\sum_{j=1}^{p} \mathrm{E}_j[\phi_n(\bar{X})] - \mathrm{E}_0[\phi_n(\bar{X})]. \tag{41}$$

Under $\mu = \mu[j]$, we have $\bar{X} \sim N(\mu[j], n^{-1}\Sigma)$, so that

$$\mathrm{E}_j[\phi(\bar{X})] = \mathrm{E}_0[e^{n\theta\bar{X}_j/\sigma_j - n\theta^2/2}\phi_n(\bar{X})].$$

Hence the right side on (41) is written as

$$\mathrm{E}_0\left[\left\{\frac{1}{p}\sum_{j=1}^{p} e^{n\theta\bar{X}_j/\sigma_j - n\theta^2/2} - 1\right\}\phi_n(\bar{X})\right] \le \mathrm{E}_0\left[\left|\frac{1}{p}\sum_{j=1}^{p} e^{n\theta\bar{X}_j/\sigma_j - n\theta^2/2} - 1\right|\right].$$

Note that under $\mu = 0$, $\sqrt{n}\bar{X}_1/\sigma_1, \dots, \sqrt{n}\bar{X}_p/\sigma_p \sim N(0,1)$ i.i.d. Hence we obtain the assertion (22).

When $\sigma_1^2, \dots, \sigma_p^2$ are unknown, the previous argument shows that the assertion (22) holds for any test with size α that depends on $X_1, \dots, X_n$ and $\sigma_1^2, \dots, \sigma_n^2$, from which we can immediately see that the assertion (22) holds for any test with size α that depends only on $X_1, \dots, X_n$.

The last assertion follows from application of Lemma 6.2 in Dümbgen and Spokoiny (2001). ☐

A.8. **Proof of Lemma 4.2.** Let $j^* \in \{1, \dots, p\}$ be any index such that $\mu_{j^*}/\sigma_{j^*} = \max_{1\le j\le p}(\mu_j/\sigma_j)$. Then whenever $\max_{1\le j\le p}|\widehat{\sigma}_j/\sigma_j - 1| \le \delta$,

$$\begin{aligned} T &\ge \sqrt{n}\widehat{\mu}_{j^*}/\widehat{\sigma}_{j^*} = \sqrt{n}\mu_{j^*}/\widehat{\sigma}_{j^*} + \sqrt{n}(\widehat{\mu}_{j^*} - \mu_{j^*})/\widehat{\sigma}_{j^*} \\ &\ge ((1-2\delta)/(1-\delta)) \cdot \sqrt{n}\mu_{j^*}/\sigma_{j^*} + \sqrt{n}(\widehat{\mu}_{j^*} - \mu_{j^*})/\widehat{\sigma}_{j^*} \\ &\ge (1 + \epsilon + \underline{\epsilon})\sqrt{2\log(p/\alpha)} + \sqrt{n}(\widehat{\mu}_{j^*} - \mu_{j^*})/\widehat{\sigma}_{j^*}, \end{aligned}$$

so that

$$\sqrt{n}(\widehat{\mu}_{j^*} - \mu_{j^*})/\widehat{\sigma}_{j^*} > -\epsilon\sqrt{2\log(p/\alpha)} \Rightarrow T > \widehat{c}(\alpha).$$

Hence we have

$$\begin{aligned} &\mathrm{P}(T > \widehat{c}(\alpha)) \\ &\ge \mathrm{P}\left(\{T > \widehat{c}(\alpha)\} \cap \left\{\max_{1\le j\le p}|\widehat{\sigma}_j/\sigma_j - 1| \le \delta\right\}\right) \\ &\ge \mathrm{P}\left(\left\{\sqrt{n}(\widehat{\mu}_{j^*} - \mu_{j^*})/\widehat{\sigma}_{j^*} > -\epsilon\sqrt{2\log(p/\alpha)}\right\} \cap \left\{\max_{1\le j\le p}|\widehat{\sigma}_j/\sigma_j - 1| \le \delta\right\}\right) \\ &\ge \mathrm{P}\left(\left\{\sqrt{n}(\widehat{\mu}_{j^*} - \mu_{j^*})/\sigma_{j^*} > -(1-\delta)\epsilon\sqrt{2\log(p/\alpha)}\right\} \cap \left\{\max_{1\le j\le p}|\widehat{\sigma}_j/\sigma_j - 1| \le \delta\right\}\right) \\ &\ge \mathrm{P}\left(\sqrt{n}(\widehat{\mu}_{j^*} - \mu_{j^*})/\sigma_{j^*} > -(1-\delta)\epsilon\sqrt{2\log(p/\alpha)}\right) - \mathrm{P}(\max_{1\le j\le p}|\widehat{\sigma}_j/\sigma_j - 1| > \delta). \end{aligned}$$

By Markov's inequality, we have

$$\begin{aligned}
&\mathrm{P}\left(\sqrt{n}(\widehat{\mu}_{j^*} - \mu_{j^*})/\sigma_{j^*} > -(1-\delta)\epsilon\sqrt{2\log(p/\alpha)}\right) \\
&= 1 - \mathrm{P}\left(\sqrt{n}(\mu_{j^*} - \widehat{\mu}_{j^*})/\sigma_{j^*} > (1-\delta)\epsilon\sqrt{2\log(p/\alpha)}\right) \\
&\geq 1 - \frac{1}{2(1-\delta)^2\epsilon^2\log(p/\alpha)}.
\end{aligned}$$

This completes the proof. □

A.9. **Proof of Corollary 4.1.** Here c, C denote generic positive constants depending only on α, c_1, C_1; their values may change from place to place. We begin with noting that since $B_n^2 \log^{3/2} p \leq C_1 n^{1/2-c_1}$, it follows from Lemma A.5 that there exists $\delta_n < \min\{C\log^{-1/2} p, 1/2\}$ such that

$$\mathrm{P}\left(\max_{1\leq j\leq p}|(\widehat{\sigma}_j/\sigma_j) - 1| > \delta_n\right) \leq Cn^{-c}.$$

Hence, by Lemma 4.2, we only have to verify that with probability one,

$$\widehat{c}(\alpha) \leq (1 + C\log^{-1/2} p)\sqrt{2\log(p/\alpha)}. \tag{42}$$

To this end, since $\beta_n \leq \alpha/3$, we note that

$$c^{SN,MS}(\alpha) \leq c^{SN}(\alpha/3), \ c^{MB,MS}(\alpha) \vee c^{HB}(\alpha) \leq c^{MB}(\alpha/3),$$

so that it suffices to verify (42) with $\widehat{c}(\alpha) = c^{SN}(\alpha)$ and $c^{MB}(\alpha)$.

For $\widehat{c}(\alpha) = c^{SN}(\alpha)$, since $\Phi^{-1}(1-p/\alpha) \leq \sqrt{2\log(p/\alpha)}$ and $\log^{3/2} p \leq Cn$ (recall that $B_n \geq 1$), it is straightforward to see that (42) is verified. For $\widehat{c}(\alpha) = c^{MB}(\alpha)$, it follows from Lemma A.4 that $c^{MB}(\alpha) \leq \sqrt{2\log p} + \sqrt{2\log(1/\alpha)}$, so that (42) is verified. This completes the proof. □

A.10. **Proof of Theorem 5.1.** The theorem readily follows from Theorems 3.1-3.5. □

A.11. **Proof of Theorem 6.1.** Here as before c, C denote generic positive constants depending only on c_1, C_1; their values may change from place to place. Let $Z^*_{ij} = (X^*_{ij} - \widehat{\mu}_j)/\widehat{\sigma}_j$, and

$$W^*_0 = \max_{1\leq j\leq p}\sqrt{n}\mathbb{E}_n[Z^*_{ij}].$$

Define the event A'_n by

$$A'_n = \{|Z_{ij}| \leq D_n, \ 1 \leq \forall i \leq n, 1 \leq \forall j \leq p\} \cap \{|\widehat{\sigma}_j/\sigma_j - 1| \leq 1/2, \ 1 \leq \forall j \leq p\}.$$

By Lemma A.5, we have $\mathrm{P}(A'_n) > 1 - Cn^{-c}$. On the event A_n,

$$\begin{aligned}
\max_{1\leq i\leq n}\max_{1\leq j\leq p}|Z^*_{ij}| &\leq \max_{1\leq i\leq n}\max_{1\leq j\leq p}\frac{|X_{ij} - \widehat{\mu}_j|}{\widehat{\sigma}_j} \\
&\leq \max_{1\leq i\leq n}\max_{1\leq j\leq p}\frac{\sigma_j}{\widehat{\sigma}_j}|Z_{ij} - \mathbb{E}_n[Z_{ij}]| \leq (3/2)\cdot 2D_n = 3D_n,
\end{aligned}$$

so that, under the present assumption that $D_n^2 \log^{7/2}(pn) \le C_1 n^{1/2-c_1}$, by Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2013a), we have

$$\sup_{t\in\mathbb{R}} |\mathrm{P}(W_0^* \le t \mid X_1^n) - \mathrm{P}(W \le t \mid X_1^n)| \le Cn^{-c}, \tag{43}$$

where W is defined in (15) (note that conditional on X_1^n, $\sqrt{n}\mathbb{E}_n[Z_i^*]$ has mean zero and the same covariance matrix as $(\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \widehat{\mu}_j)]/\widehat{\sigma}_j)_{1\le j\le p}$).

Moreover, observe that

$$|W^* - W_0^*| \le \max_{1\le j\le p} |(\widehat{\sigma}_j/\widehat{\sigma}_j^*) - 1| \times |W_0^*| =: I + II.$$

On the event A_n', by Lemma A.5 (applied with $Z_{ij} = Z_{ij}^*$) together with a simple argument, we have

$$\mathrm{P}(I > Cn^{-1/2+c_1/8} D_n^2 \log p \mid X_1^n) \le Cn^{-c_1/8}.$$

On the same event, by Lemma A.3, we have

$$\mathrm{E}[II \mid X_1^n] \le C(\sqrt{\log p} + D_n \log p/\sqrt{n}) \le C\sqrt{\log p},$$

so that by Markov's inequality, $\mathrm{P}(II > n^{c_1/8}\sqrt{\log p} \mid X_1^n) \le Cn^{-c_1/8}$. Hence we conclude that, on the event A_n',

$$\mathrm{P}(|W^* - W_0^*| > \zeta_n \mid X_1^n) \le Cn^{-c}, \tag{44}$$

where $\zeta_n = Cn^{-1/2+c_1/4} D_n \log^{3/2} p$. Here note that

$$\zeta_n \sqrt{\log p} \le Cn^{-c}. \tag{45}$$

Combining (43), (44), (45) and Lemma A.4 (ii), we can deduce that, on the event A_n',

$$\sup_{t\in\mathbb{R}} |\mathrm{P}(W^* \le t \mid X_1^n) - \mathrm{P}(W \le t \mid X_1^n)| < Cn^{-c}.$$

[Note here that conditional on X_1^n, each $\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \widehat{\mu}_j)]/\widehat{\sigma}_j$ is a standard normal random variable.] This leads to

$$c^{MB}(\alpha + \nu_n) \le c^{EB}(\alpha) \le c^{MB}(\alpha - \nu_n),$$

with $\nu_n = Cn^{-c}$. Therefore, all the conclusions of the theorem follow from Theorem 3.3. This completes the proof. □

References

Andrews, D. and Jia-Barwick, P. (2012). Inference for parameters defined by moment inequalities: a recommended moment selection procedure. *Econometrica* **80** 2805-2826.

Andrews, D. and Guggenberger, P. (2009). Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities. *Econometric Theory* **25** 669-709.

Andrews, D. and Shi, X. (2013). Inference based on conditional moment inequalities. *Econometrica* **81** 609-666.

Andrews, D. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78** 119-157.

Armstrong, T. (2011). Asymptotically exact inference in conditional moment inequality models. arXiv:1112.1024.

Armstrong, T. and Chan, H. (2012). Multiscale adaptive inference on conditional moment inequalities. arXiv:1212.5729.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica* **6** 311-329.

Bajari, P., Benkard, C., and Levin, J. (2007). Estimating dynamic models of imperfect competition. *Econometrica* **75** 1331-1370.

Bentkus, V. (2003). On the dependence of the Berry-Esseen bound on dimension. *J. Statist. Plann. Infer.* **113** 385-402.

Beresteanu, A., Molchanov, I., and Molinari, F. (2011). Sharp identification regions in models with convex moment predictions. *Econometrica* **79** 1785-1821.

Bugni, F. (2011). A comparison of inferential methods in partially identified models in terms of error in the coverage probability. Preprint.

Canay, I. (2010). EL inference for partially identified models: large deviations optimality and bootstrap validity. *J. Econometrics* **156** 408-425.

Chazal, F., Fasy, B., Lecci, F., Rinaldo, A., and Wasserman, L. (2013). Stochastic convergence of persistence landscapes and silhouettes. arXiv:1312.0308.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013a). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786-2819.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013b). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. arXiv:1301.4807.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013c). Anti-concentration and honest, adaptive confidence bands. arXiv:1303.7152.

Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica* **75** 1243-1248.

Chernozhukov, V., Lee, S., and Rosen, A. (2013). Intersection bounds: estimation and inference. *Econometrica* **81** 667-737.

Chesher, A., Rosen, A., and Smolinski, K. (2013). An instrumental variable model of multiple discrete choice. *Quantitative Economics* **4** 157-196.

Chetverikov, D. (2011). Adaptive test of conditional moment inequalities. arXiv:1201.0167.

Chetverikov, D. (2012). Testing regression monotonicity in econometric models. arXiv:1212.6757.

Ciliberto, F. and Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica* **77** 1791-1828.

Dudley, R.M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press.

Dümbgen, L. and Spokoiny, V.G. (2001). Multiscale testing of qualitative hypothesis. *Ann. Statist.* **29** 124-152.

Einmahl, U. and Li, D. (2008). Characterization of LIL behavior in Banach space. *Trans. Amer. Math. Soc.* **360** 6677-6693.

Galichon, A. and Henry, M. (2011). Set identification in models with multiple equilibria. *Rev. Econ. Stud.* **78** 1264-1298.

Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724-739.

Jing, B., Shao, Q., and Wang, Q. (2003). Self-normalized Cramer-type large deviations for independent random variables. *Ann. Prob.* **31** 2167-2215.

Lai, T.L., de la Peña, V., and Shao, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications.* Springer.

Lee, S., Song, K., and Whang, Y.-J. (2013a). Testing functional inequalities. *J. Econometrics* **172** 14-32.

Lee, S., Song, K., and Whang, Y.-J. (2013b). Testing for a general class of functional inequalities. arXiv:1311.1595.

Li, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001-1008.

Menzel, K. (2008). Consistent estimation with many moment inequalities. Preprint.

Pakes, A. (2010). Alternative models for moment inequalities. *Econometrica* **78** 1783-1822.

Praestgaard, J. and Wellner, J.A. (1993). Exchangeably weighted bootstraps of the general empirical processes. *Ann. Probab.* **21** 2053-2086.

Romano, J. and Shaikh, A. (2008). Inference for identifiable parameters in partially identified econometric models. *J. Statist. Plann. Infer.* **139** 2786-2807.

Romano, J., Shaikh, A., and Wolf, M. (2013). A practical two-step method for testing moment inequalities with an application to inference in partially identified models. Preprint.

Romano, J. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica* **73** 1237-1282.

Rosen, A. (2008). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *J. Econometrics* **146** 107-117.

Talagrand, M. (2003). *Spin Glasses: A Challenge for Mathematicians.* Springer.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer.

Wasserman, L., Kolar, M. and Rinaldo, A. (2013). Estimating undirected graphs under weak assumptions. arXiv:1309.6933.

Wolak, F. (1991). The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrica* **59** 981-995.

(V. Chernozhukov) Department of Economics and Operations Research Center, MIT, 50 Memorial Drive, Cambridge, MA 02142, USA.
E-mail address: vchern@mit.edu

(D. Chetverikov) Department of Economics, UCLA, 315 Portola Plaza, Bunche Hall Room 8283, Los Angeles, CA 90024, USA.
E-mail address: chetverikov@econ.ucla.edu

(K. Kato) Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
E-mail address: kkato@e.u-tokyo.ac.jp