Testing Many Moment Inequalities

Victor Chernozhukov
Denis Chetverikov
Kengo Kato

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP52/14

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 a variety of economic applications where the problem of testing many moment inequalities appears; a notable example is a market structure 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 statistic. Specifically, we consider critical values based upon (i) the union bound combined with a moderate deviation inequality for self-normalized sums, (ii) the multiplier and empirical bootstraps, and (iii) two-step and three-step variants of (i) and (ii) by incorporating selection of uninformative inequalities that are far from being binding and novel selection of weakly informative inequalities that are potentially binding but do not provide first order information. 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 large classes of underlying distributions. Moreover, when p grows with n, we show that all of our tests are (minimax) optimal in the sense that they are uniformly consistent against alternatives whose "distance" from the null is larger than the threshold $(2(\log p)/n)^{1/2}$, while *any* test can only have trivial power in the worst case when the distance is smaller than the threshold. Finally, we show validity of a test based on block multiplier bootstrap in the case of dependent data under some general mixing conditions.

Date: First public version: December 2013 (arXiv:1312.7614v1). This version: December 16, 2014.

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

We are grateful to Jin Hahn, 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 and USC. 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.

1. Introduction

In recent years, the moment inequalities framework has developed into a powerful tool for analyzing partially identified models. Many papers studied models with a finite and fixed (and so asymptotically small) number of both conditional and unconditional moment inequalities; see the list of references below. In practice, however, the number of moment inequalities implied by the model is often large. For example, one of the main classes of partially identified models arise from problems of estimating games with multiple equilibria, and even relatively simple static games typically produce a large set of moment inequalities; see, for example, Theorem 1 in Galichon and Henry (2011). More complicated dynamic models, including dynamic games of imperfect information, produce even larger sets of moment inequalities. Researchers therefore had to rely on ad hoc, case-specific, arguments to select a small subset of moment inequalities to which the methods available in the literature so far could be applied. In this paper, we develop systematic methods to treat *many* moment inequalities. Our methods are universally applicable in any setting leading to many moment inequalities.[1]

There are variety of economic applications where the problem of testing many moment inequalities appears. One example is the discrete choice 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. Yet 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 studied in Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011), Chesher, Rosen, and Smolinski

[1]In some special settings, such as those studied in Theorem 4 of Galichon and Henry (2011), the number of moment inequalities can be dramatically reduced without blowing up the identified set (and so without any subjective choice). However, there are no theoretically justified procedures that would generically allow to decrease the number of moment inequalities in all settings.

In addition, it is important to note that in practice, it may be preferable to use more inequalities than those needed for sharp identification of the model. Indeed, selecting inequalities for statistical inference and selecting a minimal set of inequalities that suffice for sharp identification are rather different problems since the latter problem relies upon the knowledge of the inequalities and does not take into account the noise associated with estimation of inequalities. For example, if a redundant inequality can be estimated with high precision, it may be beneficial to use it for inference in addition to inequalities needed for sharp identification since such an inequality may improve finite sample statistical properties of the inferential procedure.

(2013), and Chesher and Rosen (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.[2] This means that the existing inference methods for conditional moment inequalities, albeit fruitful in many cases, do not apply to 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.

There are also many empirical studies where many moment inequalities framework could be useful. Among others, these are Ciliberto and Tamer (2009) who estimated the empirical importance of firm heterogeneity as a determinant of market structure in the US airline industry,[3] Holmes (2011) who estimated the dynamic model of Wal-Mart expansion,[4] and Ryan (2012) who estimated the welfare costs of the 1990 Amendments to the Clean Air Act on the U.S. Portland cement industry.[5]

To formally describe the problem, let $X_1, \dots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random vectors in $\mathbb{R}^p$, where

[2]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.

[3]Ciliberto and Tamer (2009) had 2742 markets and used four major airline companies and two aggregates of medium size and low cost companies that lead to $2^{4+2+1} = 128$ moment inequalities. However, as established in Theorem 1 of Galichon and Henry (2011), sharp identification bounds in the Ciliberto and Tamer model would require around $2^{2^{4+2}} = 2^{64}$ inequalities. In addition, using techniques developed in this paper, it would be possible to estimate a more detailed model, with a larger set of airline companies taken individually and using the Ciliberto and Tamer type inequalities.

[4]Holmes (2011) derived moment inequalities from ruling out deviations from the observed Wal-Mart behavior as being suboptimal. He considered the set of potential deviations where the opening dates of some Wal-Mart stores are reordered, and explicitly acknowledged that this leads to the enormous number of inequalities (in fact, this is a number of permutations of 3176 Wal-Mart stores, up-to a restriction that the stores opened in the same year can not be permuted). Therefore, he restricted attention to deviations consisting of pairwise resequencing where each deviation switches the opening dates of only two stores. However, one could argue that deviations in the form of block resequencing where the opening dates of blocks of stores are switched are also informative since one of the main features of the Wal-Mart strategy is to pack stores closely together, so that it is easy to set up a distribution network and save on trucking costs.

[5]Ryan (2012) adapted an estimation strategy proposed in Bajari, Benkard, and Levin (2007). He had 517 market-year observations and considered 1250 alternative policies to generate a set of inequalities.

$X_i = (X_{i1}, \dots, X_{ip})^T$. For $1 \leq j \leq p$, write $\mu_j := \mathrm{E}[X_{1j}]$. We are interested in testing the null hypothesis

$$H_0 : \mu_j \leq 0 \quad \text{for all } j = 1, \dots, p, \tag{1}$$

against the alternative

$$H_1 : \mu_j > 0 \quad \text{for some } j = 1, \dots, p. \tag{2}$$

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. The primal feature of this paper is that the number of moment inequalities p is allowed to be larger or even much larger than the sample size n.

We consider the test statistic given by the maximum over p Studentized (or t-type) statistics (see (13) 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 bound combined with a moderate deviation inequality for self-normalized sums, and (ii) bootstrap methods. We will call the first option the *SN method* (SN refers to the abbreviation of "Self-Normalized"). Among bootstrap methods, we consider multiplier and empirical bootstrap procedures abbreviated as *MB and EB methods*. The SN method is analytical and is very easy to implement. As such, the SN method is particularly useful for grid search when the researcher is interested in constructing the confidence region for the identified set in the parametric model defined via moment inequalities as in Section 6. Bootstrap methods are simulation-based and computationally harder.[6] However, an important feature of bootstrap methods is that they take into account correlation structure of the data and yield lower critical values leading to more powerful tests than those obtained via the SN method. In particular, if the researcher incidentally repeated the same inequality twice or, more importantly, included inequalities with very similar informational content (that is, highly correlated inequalities), the MB/EB methods would be able to account of this and would automatically disregard or nearly disregard these duplicated or nearly duplicated inequalities, without inflating the critical value.

We also consider two-step methods by incorporating inequality selection procedures. The two-step methods get rid of most of *uninformative* inequalities, that is inequalities j with $\mu_j < 0$ if μ_j is not too close to 0. By dropping the uninformative inequalities, the two-step methods produce more powerful tests than those based on one-step methods, that is, methods without inequality selection procedures.

Moreover, we develop novel three-step methods by incorporating double inequality selection procedures. The three-step methods are suitable in

[6] In fact, the MB and EB methods are computationally also rather simple. For example, it took us only about 2 hours to conduct all the Monte Carlo experiments described in Section 8, which is remarkably small time for Monte Carlo experiments in our experience.

parametric models defined via moment inequalities and allow to drop *weakly informative* inequalities in addition to uninformative inequalities.[7] Specifically, consider the model $\mathrm{E}[g_j(\xi,\theta)] \leq 0$ for all $j = 1,\ldots,p$ where ξ is a vector of random variables, θ a vector of parameters, and $g_1,\ldots,g_p$ a set of functions. Suppose that the researcher is interested in testing the null hypothesis $\theta = \theta_0$ against the alternative $\theta \neq \theta_0$ based on the i.i.d. data $\xi_1,\ldots,\xi_n$, so that the problem reduces to (1)-(2) by setting $X_{ij} = g_j(\xi_i,\theta_0)$. We say that the inequality j is weakly informative if the function $\theta \mapsto \mathrm{E}[g_j(\xi,\theta)]$ is flat or nearly flat at $\theta = \theta_0$. Dropping weakly informative inequalities allows us to derive tests with higher local power since these inequalities can only provide a weak signal of violation of the null hypothesis when θ is close to θ_0.

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; indeed, p can be of order $\exp(n^c)$ for some $c > 0$. In addition, we emphasize that although we are primarily interested in the case with p (much) larger than n, our methods remain valid when p is small or comparable to n.

We also show optimality of our tests from a minimax point of view.[8] Specifically, we consider the alternative hypotheses whose "distance" from the null is $r > 0$ in the sense that $H_{1,r} : \max_{1\leq j\leq p}(\mu_j/\sigma_j) \geq r$, where $\sigma_j^2 := \mathrm{Var}(X_{1j})$. Intuitively, the smaller $r > 0$ is, the more difficult to distinguish $H_{1,r}$ from H_0 is. We show that, when $p = p_n \to \infty$ as $n \to \infty$, *any* test can only have trivial power against $H_{1,\underline{r}_n}$ in the worst case where $\underline{r}_n = (1-\epsilon_n)\sqrt{2(\log p_n)/n}$ and where ϵ_n is any positive sequence with $\epsilon_n \to 0$ sufficiently slowly, while all of our tests described above are uniformly consistent against $H_{1,\overline{r}_n}$ where $\overline{r}_n = (1+\epsilon_n)\sqrt{2(\log p_n)/n}$.

An important feature of our methods is that increasing the set of moment inequalities has no or little effect on the critical value. In particular, our critical values are always bounded by a slowly varying function $(\log p)^{1/2}$ (up-to a constant). This implies that instead of making a subjective choice of inequalities, the researcher can use all (or at least a large set of) available inequalities, and the results will be not-too-different from those based only on the inequalities that violate H_0 (the latter procedure is of course infeasible, and the slow growth of the critical value can be thought of as a small

[7]The same methods can be extended to nonparametric models as well. In this case, θ appearing below in this paragraph should be considered as a sieve parameter.

[8]See, for example, Ingster (1993) and Ingster and Suslina (2003) for the minimax approach to hypothesis testing in nonparametric statistical models.

cost of data-driven adaptation to the inequalities that violate H_0). This feature of our methods is akin to that in modern high dimensional/big data techniques like Lasso and Dantzig selector that allow for variable selection in exchange for small cost in precision of model estimates; see, for example, Bickel, Ritov, and Tsybakov (2009) for analysis and discussion of methods of estimating high dimensional models.

Moreover, we consider two extensions of our results in Section 7. In the first extension, we consider testing many moment inequalities for dependent data. Specifically, we prove validity of a test based on the (block) multiplier bootstrap under β-mixing conditions; see Section 7 for the definitions. As in other parts of the paper, our results allow p to be potentially much larger than n. Our results complement the set of impressive results in Zhang and Cheng (2014) who, independently from us and around the same time, considered the case of the functionally dependent time series data (the concept of functional dependence was introduced in Wu (2005) and is different from β-mixing). Thus, both our paper and Zhang and Cheng (2014) extend Gaussian approximation and bootstrap results of Chernozhukov, Chetverikov, and Kato (2013a) to the case of dependent data but under different dependence conditions (that do not nest each other). The results obtained in these two papers are strongly complementary and, taken together, cover a wide variety of dependent data processes, thereby considerably expanding the applicability of the proposed tests. In the second extension, we allow for approximate inequalities to account of the case where an approximation error arises either from estimated nuisance parameters or from the need to linearize the inequalities. Both of these extensions are important for inference in dynamic models such as those considered in Bajari, Benkard, and Levin (2007).

The problem of testing moment inequalities described above is "dual" of that of constructing confidence regions for identifiable parameters in partially identified models where identified sets are given by 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 large classes of underlying distributions, so that the resulting confidence regions are *(asymptotically) honest* to such classes (see Section 6 for the precise meaning).

The literature on testing (unconditional) moment inequalities is large; see 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), Bajari, Benkard, and Levin (2007), and Pakes (2010). Menzel (2009) considered inference for many moment inequalities, but with p growing at most as $n^{2/7}$ (and hence p being much smaller than n). Also his approach and test statistics are different from ours. Finally, Allen (2014) recently suggested further extensions and refinements of our new methods. In particular, he noticed that the truncation threshold for our selection procedures can be taken slightly lower (in absolute value) than what we use; he studied an iterative procedure based on Chetverikov (2011); and he considered moment re-centering procedure similar to that developed in Romano, Shaikh, and Wolf (2013). The latter two possibilities were already noted in the previous versions of our paper.[9]

The remainder of the paper is organized as follows. In the next section, we discuss several motivating examples. In Section 3, we build our test statistic. In Section 4, we derive various ways of computing critical values for the test statistic, including the SN, MB, and EB methods and their two-step and three-step variants discussed above, and prove their validity. In Section 5, we show asymptotic minimax optimality of our tests. In Section 6, we present the corresponding results on construction of confidence regions for identifiable parameters in partially identified models. In Section 7, we present two extensions discussed above. All the technical proofs are deferred to the Appendix. Finally, the Supplemental Material discusses some details on the dynamic model of imperfect competition example discussed in Section 2.

1.1. **Notation and convention.** 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, we 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, and our results also apply to the case with fixed p. Finally, throughout the paper, we assume that $n \geq 2$ and $p \geq 2$.

[9]See the 2013 version of our paper at arXiv:1312.7614v1.

2. Motivating examples

In this section, we provide three examples that motivate the framework where the number of moment inequalities p is large and potentially much larger than the sample size n. In these examples, one actually has many conditional rather than unconditional moment inequalities. Therefore, we emphasize that our results cover the case of many conditional moment inequalities as well.[10] As these examples demonstrate, there is a variety of economic models leading to the problem of testing many unconditional and/or many conditional moment inequalities to which the methods available in the literature so far can not be applied, and which, therefore, requires the methods developed in this paper.

2.1. **Market structure model.** This example is based on Ciliberto and Tamer (2009). Let m denote the number of firms that could potentially enter the market. Let m-tuple $D = (D_1, \dots, D_m)$ denote entry decisions of these firms; that is, $D_j = 1$ if the firm j enters the market and $D_j = 0$ otherwise. Let $\mathcal{D}$ denote the set of possible values of D. Clearly, the number of elements d of the set $\mathcal{D}$ is $|\mathcal{D}| = 2^m$.

Let X and ε denote (exogenous) characteristics of the market as well as characteristics of the firms that are observed and not observed by the researcher, respectively. The profit of the firm j is given by

$$\pi_j(D, X, \varepsilon, \theta),$$

where the function π_j is known up to a parameter θ. Assume that both X and ε are observed by the firms and that a Nash equilibrium is played, so that for each j,

$$\pi_j((D_j, D_{-j}), X, \varepsilon, \theta) \geq \pi_j((1 - D_j, D_{-j}), X, \varepsilon, \theta),$$

where D_{-j} denotes decisions of all firms excluding the firm j. Then one can find set-valued functions $R_1(d, X, \theta)$ and $R_2(d, X, \theta)$ such that d is the unique equilibrium whenever $\varepsilon \in R_1(d, X, \theta)$, and d is one of several equilibria whenever $\varepsilon \in R_2(d, X, \theta)$. When $\varepsilon \in R_1(d, X, \theta)$ for some $d \in \mathcal{D}$, we know for sure that $D = d$ but when $\varepsilon \in R_2(d, X, \theta)$, the probability that $D =$

[10]Indeed, consider conditional moment inequalities of the form

$$\mathrm{E}[g_j(Y) \mid Z] \leq 0 \quad \text{for all } j = 1, \dots, p' \tag{3}$$

where (Y, Z) is a pair of random vectors and $g_1, \dots, g_{p'}$ is a set of functions with p' being large. Let $\mathcal{Z}$ be the support of Z and assume that $\mathcal{Z}$ is a compact set in $\mathbb{R}^l$. Then, following Andrews and Shi (2013), one can construct an infinite set $\mathcal{I}$ of instrumental functions $I : \mathcal{Z} \to \mathbb{R}$ such that $I(z) \geq 0$ for all $z \in \mathcal{Z}$ and (3) holds if and only if

$$\mathrm{E}[g_j(Y) I(Z)] \leq 0 \quad \text{for all } j = 1, \dots, p' \text{ and all } I \in \mathcal{I}.$$

In practice, one can choose a large subset $\mathcal{I}_n$ of $\mathcal{I}$ and consider testing $p = p'|\mathcal{I}_n|$ moment inequalities

$$\mathrm{E}[g_j(Y) I(Z)] \leq 0 \quad \text{for all } j = 1, \dots, p' \text{ and all } I \in \mathcal{I}_n. \tag{4}$$

If $\mathcal{I}_n$ is large enough and grows sufficiently fast with n, the test of (3) based on (4) will be consistent.

d depends on the equilibrium selection mechanism, and, without further information, can be anything in $[0,1]$. Therefore, we have the following bounds

$$\begin{aligned}\mathrm{E}\left[1\{\varepsilon \in R_1(d,X,\theta) \mid X\right] &\leq \mathrm{E}\left[1\{D=d\} \mid X\right] \\ &\leq \mathrm{E}\left[1\{\varepsilon \in R_1(d,X,\theta) \cup R_2(d,X,\theta)\} \mid X\right],\end{aligned}$$

for all $d \in \mathcal{D}$. Further, assuming that the conditional distribution of ε given X is known (alternatively, it can be assumed that this distribution is known up to a parameter that is a part of the parameter θ), both the left- and the right-hand sides of these inequalities can be calculated. Denote them by $P_1(d,X,\theta)$ and $P_2(d,X,\theta)$, respectively, to obtain

$$P_1(d,X,\theta) \leq \mathrm{E}\left[1\{D=d\} \mid X\right] \leq P_2(d,X,\theta) \text{ for all } d \in \mathcal{D}. \tag{5}$$

These inequalities can be used for inference about the parameter θ. Note that the number of inequalities in (5) is $2|\mathcal{D}| = 2^{m+1}$, which is a large number even if m is only moderately large. Moreover, these inequalities are conditional on X. For inference about the parameter θ, each of these inequalities can be transformed into a large and increasing number of unconditional inequalities as described above. Also, if the firms have more than two decisions, the number of inequalities will be even (much) larger. Therefore, our framework is exactly suitable for this example.

2.2. **Discrete choice model with endogeneity.** Our second example is based on Chesher, Rosen, and Smolinski (2013). The source of many moment inequalities in this example is different from that in the previous example. Consider an individual who is choosing an alternative d from a set $\mathcal{D}$ of available options. Let $M = |\mathcal{D}|$ denote the number of available options. Let D denote the choice of the individual. From choosing an alternative d, the individual obtains the utility

$$u(d,X,V),$$

where X is a vector of observable (by the researcher) covariates and V is a vector of unobservable (by the researcher) utility shifters. The individual observes both X and V and makes a choice based on utility maximization, so that D satisfies

$$u(D,X,V) \geq u(d,X,V) \text{ for all } d \in \mathcal{D}.$$

The object of interest in this model is the pair (u, P_V) where P_V denotes the distribution of the vector V.

In many applications, some components of X may be endogenous in the sense that they are not independent of V. Therefore, to achieve (partial) identification of the pair (u, P_V), following Chesher, Rosen, and Smolinski (2013), assume that there exists a vector Z of observable instruments that are independent of V. Let $\mathcal{V}$ denote the support of V, and let $\tau(d,X,u)$

denote the subset of $\mathcal{V}$ such that $D = d$ whenever $X = x$ and $V \in \tau(d, x, u)$, so that

$$V \in \tau(D, X, u). \tag{6}$$

Then for any set $S \subset \mathcal{V}$,

$$\mathrm{E}\left[1\{V \in S\}\right] = \mathrm{E}\left[1\{V \in S\} \mid Z\right] \geq \mathrm{E}\left[1\{\tau(D, X, u) \subset S\} \mid Z\right], \tag{7}$$

where the equality follows from independence of V from Z, and the inequality from (6). Note that the left-hand side of (7) can be calculated (for fixed distribution P_V) and equals $P_V(S)$, so that we obtain

$$P_V(S) \geq \mathrm{E}\left[1\{\tau(D, X, u) \subset S\} \mid Z\right] \text{ for all } S \in \mathcal{S}, \tag{8}$$

where $\mathcal{S}$ is some collection of sets in $\mathcal{V}$. Inequalities (8) can be used for inference about the pair (u, P_V). A natural question then is what collection of sets $\mathcal{S}$ should be used in (8). Chesher, Rosen, and Smolinski (2013) showed that sharp identification of the pair (u, P_V) is achieved by considering all unions of sets on the support of $\tau(D, X, u)$ with the property that the union of the interiors of these sets is a connected set. When X is discrete with the support consisting of m points, this implies that the class $\mathcal{S}$ may consist of $M \cdot 2^m$ sets, which is a large number even for moderately large m. Moreover, as in our previous example, inequalities in (8) are conditional giving rise to even a larger set of inequalities when transformed into unconditional ones. Therefore, our framework is again exactly suitable for this example.

Also, we note that the model described in this example fits as a special case into a Generalized Instrumental Variable framework set down and analyzed by Chesher and Rosen (2013), where the interested reader can find other examples leading to many moment inequalities.

2.3. **Dynamic model of imperfect competition.** This example is based on Bajari, Benkard, and Levin (2007). In this example, many moment inequalities arise from ruling out deviations from best responses in a dynamic game. Consider a market consisting of N firms. Each firm j makes a decision $A_{jt} \in \mathcal{A}$ at time periods $t = 0, 1, 2, \dots, \infty$. Let $A_t = (A_{1t}, \dots, A_{Nt})$ denote the N-tuple of decisions of all firms at period t. The profit of the firm j at period t, denoted by $\pi_j(A_t, S_t, \nu_{jt})$, depends on the N-tuple of decisions A_t, the state of the market $S_t \in \mathcal{S}$ at period t, and the firm- and time-specific shock $\nu_{jt} \in \mathcal{V}$. Assume that the state of the market S_t follows a Markov process, so that S_{t+1} has the distribution function $P(S_{t+1}|A_t, S_t)$, and that ν_{jt} is i.i.d. across firms j and time periods t with the distribution function $G(\nu_{jt})$. In addition, assume that when the firm j is making a decision A_{jt} at period t, it observes S_t and ν_{jt} but does not observe ν_{-jt}, the specific shocks of all its rivals, and that the objective function of the firm j at period t is to maximize

$$\mathrm{E}\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} \pi_j(A_\tau, S_\tau, \nu_{jt}) \mid S_t\right],$$

where β is a discount factor. Further, assume that a Markov Perfect Equilibrium (MPE) is played in the market. Specifically, let $\sigma_j : \mathcal{S} \times \mathcal{V} \to \mathcal{A}$ denote the MPE strategy of firm j, and let $\sigma := (\sigma_1, \dots, \sigma_N)$ denote the N-tuple of strategies of all firms. Define the value function of the firm j in the state $s \in \mathcal{S}$ given the profile of strategies σ, $V_j(s, \sigma)$, by the Belmann equation:

$$V_j(s,\sigma) := \mathrm{E}_\nu \Big[\pi_j(\sigma(s,\nu), s, \nu_j) + \beta \int V_j(s', \sigma) dP(s' \mid \sigma(s,\nu), s) \Big],$$

where $\sigma(s,\nu) = (\sigma_1(s,\nu_1), \dots, \sigma_N(s,\nu_N))$, and expectation is taken with respect to $\nu = (\nu_1, \dots, \nu_N)$ consisting of N i.i.d. random variables ν_j with the distribution function $G(\nu_j)$. Then the profile of strategies σ is an MPE if for any $j = 1, \dots, N$ and $\sigma_j' : \mathcal{S} \times \mathcal{V} \to \mathcal{A}$, we have

$$\begin{aligned} V_j(s,\sigma) &\geq V_j(s, \sigma_j', \sigma_{-j}) \\ &= \mathrm{E}_\nu \Big[\pi_j(\sigma_j'(s,\nu_i), \sigma_{-j}(s, \nu_{-j}), s, \nu_j) \\ &\qquad + \beta \int V_j(s', \sigma_j', \sigma_{-j}) dP(s' \mid \sigma_j'(s,\nu_j), \sigma_{-j}(s,\nu_{-j}), s) \Big], \end{aligned}$$

where σ_{-j} is strategies of all rivals of the firm j in the profile σ.

For estimation purposes, assume that the functions $\pi_j(A_t, S_t, \nu_{jt})$ and $G(\nu_{jt})$ are known up-to a finite dimensional parameter θ, that is we have $\pi_j(A_t, S_t, \nu_{jt}) = \pi_j(A_t, S_t, \nu_{jt}, \theta)$ and $G(\nu_{jt}) = G(\nu_{jt}, \theta)$, so that the value function $V_j(s,\sigma) = V_j(s,\sigma,\theta)$ also depends on θ, and the goal is to estimate θ. Assume that the data consist of observations on n similar markets for a short span of periods or observations on one market for n periods. In the former case, assume also that the same MPE is played in all markets.[11]

In this model, Bajari, Benkard, and Levin (2007) suggested a computationally tractable two stage procedure to estimate the structural parameter θ. An important feature of their procedure is that it does not require point identification of the model. The first stage of their procedure consists of estimating transition probability function $P(S_{t+1} | S_t, A_t)$ and policy functions (strategies) $\sigma_j(s, \nu_j)$. Following their presentation, assume that these functions are known up-to a finite dimensional parameter α, that is $P(S_{t+1} | S_t, A_t) = P(S_{t+1} | S_t, A_t, \alpha)$ and $\sigma_j(s,\nu_j) = \sigma_j(s, \nu_j, \alpha)$, and that the first stage yields a $\sqrt{n}$-consistent estimator $\widehat{\alpha}_n$ of α. Using $\widehat{\alpha}_n$, one can estimate transition probability function by $P(S_{t+1} | S_t, A_t, \widehat{\alpha}_n)$, and then one can calculate the (estimated) value function of the firm j at every state $s \in \mathcal{S}$, $\widehat{V}_j(s, \sigma', \theta)$, for any profile of strategies σ' and any value of the parameter θ using forward simulation as described in Bajari, Benkard, and Levin (2007).

[11]In the case of data consisting of observations on one market for n periods, one has to use techniques for dependent data developed in Section 7.1 of this paper. It is also conceptually straightforward to extend our techniques to the case when the data consist of observations on many markets for many periods, as happens in some empirical studies. We leave this extension for future work.

Here we have $\widehat{V}_j(s,\sigma',\theta)$ instead of $V_j(s,\sigma',\theta)$ because forward simulations are based on the estimated transition probability function $P(S_{t+1}|S_t,A_t,\widehat{\alpha}_n)$ instead of the true functions $P(S_{t+1}|S_t,A_t,\alpha)$. Then, on the second stage, one can test equilibrium conditions

$$V_j(s,\sigma_j,\sigma_{-j},\theta)\geq V_j(s,\sigma_j',\sigma_{-j},\theta)$$

for all $j=1,\dots,N$, $s\in\mathcal{S}$, and $\sigma_j'\in\Sigma$ for some set of strategies Σ by considering inequalities

$$\widehat{V}_j(s,\sigma_j(\widehat{\alpha}_n),\sigma_{-j}(\widehat{\alpha}_n),\theta)\geq \widehat{V}_j(s,\sigma_j',\sigma_{-j}(\widehat{\alpha}_n),\theta) \tag{9}$$

where $\sigma_j(\widehat{\alpha}_n)$ and $\sigma_{-j}(\widehat{\alpha}_n)$ are estimated policy functions for the firm j and all of its rivals, respectively. Inequalities (9) can be used to test hypotheses about parameter θ. The number of inequalities is determined by the number of elements in Σ. Assuming that $\mathcal{A}$, $\mathcal{S}$, and $\mathcal{V}$ are all finite, we obtain $|\Sigma|=|\mathcal{A}|^{|\mathcal{S}|\cdot|\mathcal{V}|}$, so that the total number of inequalities is $N\cdot|\mathcal{S}|\cdot|\Sigma|$, which is a very large number in all but trivial empirical applications.

Inequalities (9) do not fit directly into our testing framework (1)-(2). One possibility to go around this problem is to use a jackknife procedure. In a nutshell, assuming that the data consist of observations on n i.i.d markets, the procedure is as follows. Let $\widehat{\alpha}_n^{-i}$, $\widehat{V}_j^{-i}(s,\sigma_j(\widehat{\alpha}_n^{-i}),\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta)$, and $\widehat{V}_j^{-i}(s,\sigma_j',\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta)$ denote the leave-market-i-out estimates of α, $V_j(s,\sigma_j(\alpha),\sigma_{-j}(\alpha),\theta)$, and $V_j(s,\sigma_j',\sigma_{-j}(\alpha),\theta)$, respectively. Define

$$\begin{aligned}\widetilde{X}_{ij}(s,\theta):=\ &n\widehat{V}_j(s,\sigma_j(\widehat{\alpha}_n),\sigma_{-j}(\widehat{\alpha}_n),\theta)\\ &-(n-1)\widehat{V}_j^{-i}(s,\sigma_j(\widehat{\alpha}_n^{-i}),\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta)\end{aligned}$$

and

$$\begin{aligned}\widetilde{X}_{ij}'(s,\sigma_j',\theta):=\ &n\widehat{V}_j(s,\sigma_j',\sigma_{-j}(\widehat{\alpha}_n),\theta)\\ &-(n-1)\widehat{V}_j^{-i}(s,\sigma_j',\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta).\end{aligned}$$

Also, define

$$\widehat{X}_{ij}(s,\sigma_j',\theta):=\widetilde{X}_{ij}'(s,\sigma_j',\theta)-\widetilde{X}_{ij}(s,\theta).$$

Then under some regularity conditions including smoothness of the value function $V_j(s,\sigma)$, one can show that

$$\widehat{X}_{ij}(s,\sigma_j',\theta)=X_{ij}(s,\sigma_j',\theta)+O_P(n^{-1/2}) \tag{10}$$

where

$$\mathrm{E}[X_{ij}(s,\sigma_j',\theta)]=V_j(s,\sigma_j',\sigma_{-j},\theta)-V_j(s,\sigma,\theta) \tag{11}$$

and $X_{ij}(s,\sigma_j',\theta)$'s are independent across markets $i=1,\dots,n$. We provide some details on the derivation of (10) in the Supplemental Material. Now we can use results of Section 7.2 on testing approximate moment inequalities to do inference about the parameter θ if we replace $X_{ij}(s,\sigma_j',\theta)$ by the "data" $\widehat{X}_{ij}(s,\sigma_j',\theta)$ and we use $\sqrt{n}(\widehat{V}_j(s,\sigma_j',\sigma_{-j}(\widehat{\alpha}_n),\theta)-\widehat{V}_j(s,\sigma_j(\widehat{\alpha}_n),\sigma_{-j}(\widehat{\alpha}_n),\theta)$ instead of $\sqrt{n}\widehat{\mu}_j=n^{-1/2}\sum_{i=1}^n\widehat{X}_{ij}(s,\sigma_j',\theta)$ in the numerator of our test statistic defined in (13). Thus, this example fits into our framework as well.[12]

3. Test statistic

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

$$\mathrm{E}[X_{1j}^2] < \infty, \ \sigma_j^2 := \mathrm{Var}(X_{1j}) > 0, \ j = 1, \dots, p. \tag{12}$$

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

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

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

There are several different statistics that can be used for testing the null hypothesis (1) against the alternative (2). 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 $\mathrm{E}[|X_{1j}|^3]$ (see Section 4.1.1). As far as MB/EB 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.

[12]The jackknife procedure described above may be computationally intensive in some applications but, on the other hand, the required computations are rather straightforward. In addition, this procedure only involves the first stage estimation, which is typically computationally simple. Moreover, bootstrap procedures developed in this paper do not interact with the jackknife procedure, so that the latter procedure has to be performed only once.

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

$$T = \max_{1 \le j \le p} \frac{\sqrt{n}\widehat{\mu}_j}{\widehat{\sigma}_j}. \tag{13}$$

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

$$T > c \Rightarrow \text{reject } H_0, \tag{14}$$

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 $j = 1, \dots, p$. In such cases, we interpret the meaning of "$T > c$" in (14) as $\sqrt{n}\widehat{\mu}_j > c\widehat{\sigma}_j$ for some $j = 1, \dots, p$, which makes sense even if $\widehat{\sigma}_j^2 = 0$ for some $j = 1, \dots, p$. 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

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

The statistic T' has an advantage that it is less sensitive to outliers. However, T' leads to good power only if many inequalities 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. When p is large, as in our motivating examples, the statistic T seems preferable over T' because the critical value for the test based on T grows very slowly with p (at most as $(\log p)^{1/2}$) whereas one can expect that the critical value for the test based on T' grows at least polynomially with p.

Another alternative is the test statistic of the form

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

where $\widehat{\mu} = (\widehat{\mu}_1, \dots, \widehat{\mu}_p)^T$, $t = (t_1, \dots, t_p)^T \le 0$ means $t_j \le 0$ for all $j = 1, \dots, p$, 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.

4. 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,

$$T \le \max_{1 \le j \le p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j, \tag{15}$$

where the equality holds when all the moment inequalities are binding, that is, $\mu_j = 0$ for all $j = 1, \dots, p$. 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 $\max_{1 \le j \le p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j$. We consider two approaches to construct such critical values: self-normalized and bootstrap methods. We also consider two- and three-step variants of the methods by incorporating inequality selection.

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

$$Z_{ij} = (X_{ij} - \mu_j)/\sigma_j, \text{ and } Z_i = (Z_{i1}, \dots, Z_{ip})^T. \tag{16}$$

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

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

Note that by Jensen's inequality, $B_n \ge M_{n,4} \ge M_{n,3} \ge 1$.

4.1. Self-Normalized methods.

4.1.1. *One-step method.* The self-normalized method (abbreviated as the SN method in what follows) we consider is based upon the union bound combined with a moderate deviation inequality for self-normalized sums. Because of inequality (15), 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 may look 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 inequality 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 bound, we will show in Section 5 that the resulting test is asymptotically minimax optimal when $p = p_n \to \infty$ as $n \to \infty$.

For $j = 1, \dots, p$, define

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

By 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-hand side is increasing in U_j as long as $U_j \geq 0$. Hence under H_0,

$$\mathrm{P}(T > c) \leq \sum_{j=1}^{p} \mathrm{P}\left(U_j > c/\sqrt{1 + c^2/n}\right), \ c \geq 0. \tag{17}$$

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

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

even if Z_{ij} only have $2 + \delta$ finite moments for some $\delta > 0$. Therefore, we take the critical value as

$$c^{SN}(\alpha) = \frac{\Phi^{-1}(1 - \alpha/p)}{\sqrt{1 - \Phi^{-1}(1 - \alpha/p)^2/n}}, \tag{18}$$

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

$$\mathrm{P}(T > c^{SN}(\alpha)) \leq \alpha \left[1 + K n^{-1/2} M_{n,3}^3 \{1 + \Phi^{-1}(1 - \alpha/p)\}^3\right], \tag{19}$$

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

$$M_{n,3}^3 \log^{3/2}(p/\alpha) \leq C_1 n^{1/2 - c_1}, \tag{20}$$

then under H_0,

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

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 (12) and (20) are verified.*

Comment 4.1. The theorem assumes that $\max_{1\le j\le p} \mathrm{E}[|X_{1j}|^3] < \infty$ (so that $M_{n,3} < \infty$) but allows this quantity to diverges as $n \to \infty$ (recall $p = p_n$). In principle, $M_{n,3}$ that appears in the theorem could be replaced by $\max_{1\le j\le p}(\mathrm{E}[|Z_{1j}|^{2+\nu}])^{1/(2+\nu)}$ for $0 < \nu \le 1$, which would further weaken moment conditions; however, for the sake of simplicity of presentation, we do not explore this generalization.

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

Observe that when $\mu_j < 0$ for some $j = 1,\dots,p$, inequality (15) becomes strict, so that when there are many j for which μ_j are negative and large in absolute value, the resulting test with one-step SN critical values would tend to be unnecessarily 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 inequality selection.

More formally, let $0 < \beta_n < \alpha/3$ 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

$$\widehat{J}_{SN} := \left\{ j \in \{1,\dots,p\} : \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j > -2c^{SN}(\beta_n) \right\}. \qquad (21)$$

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

$$c^{SN,2S}(\alpha) = \begin{cases} \frac{\Phi^{-1}(1-(\alpha-2\beta_n)/\widehat{k})}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta_n)/\widehat{k})^2/n}}, & \text{if } \widehat{k} \ge 1, \\ 0, & \text{if } \widehat{k} = 0. \end{cases} \qquad (22)$$

The following theorem establishes validity of this critical value.

Theorem 4.2 (Validity of two-step SN method)**.** *Suppose that* $\sup_{n\ge 1}\beta_n \le \alpha/3$ *and there exist constants* $0 < c_1 < 1/2$ *and* $C_1 > 0$ *such that*

$$M_{n,3}^3 \log^{3/2}(p/\beta_n) \le C_1 n^{1/2-c_1}, \text{ and } B_n^2 \log^2(p/\beta_n) \le C_1 n^{1/2-c_1}. \qquad (23)$$

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

$$\mathrm{P}(T > c^{SN,2S}(\alpha)) \le \alpha + Cn^{-c}.$$

Moreover, this bound holds uniformly with respect to the common distribution of X_i *for which (12) and (23) are verified.*

4.2. **Bootstrap methods.** In this section, we consider bootstrap methods for calculating critical values. Specifically, we consider Multiplier Bootstrap (MB) and Empirical (nonparametric, or Efron's) Bootstrap (EB) methods. The methods studied 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.

4.2.1. *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 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 $Z_i = (Z_{i1}, \dots, Z_{ip})^T$ are defined in (16). *When p is fixed*, the central limit theorem guarantees that as $n \to \infty$,

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

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 $\max_{1 \leq j \leq p} \sqrt{n}\mathbb{E}_n[Z_{ij}]$ and $\max_{1 \leq j \leq p} Y_j$ 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, 2014b) shows that, under mild regularity conditions, the distribution of $\max_{1 \leq j \leq p} \sqrt{n}\mathbb{E}_n[Z_{ij}]$ can be approximated by that of $\max_{1 \leq j \leq p} Y_j$ in the sense of Kolmogorov distance *even when p is larger or much larger than n.*[13] 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.[14]

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 one the following two bootstrap procedures:

Algorithm (Multiplier bootstrap)**.**

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

2. Construct the multiplier bootstrap test statistic
$$W^{MB}=\max_{1\le j\le p}\frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_j)]}{\widehat{\sigma}_j}. \tag{24}$$

3. Calculate $c^{MB}(\alpha)$ as
$$c^{MB}(\alpha)=\text{conditional } (1-\alpha)\text{-quantile of } W^{MB} \text{ given } X_1^n. \tag{25}$$

Algorithm (Empirical bootstrap)**.**

1. Generate a bootstrap sample $X_1^*,\dots,X_n^*$ as i.i.d. draws from the empirical distribution of $X_1^n=\{X_1,\dots,X_n\}$.

2. Construct the empirical bootstrap test statistic
$$W^{EB}=\max_{1\le j\le p}\frac{\sqrt{n}\mathbb{E}_n[X_{ij}^*-\widehat{\mu}_j]}{\widehat{\sigma}_j}. \tag{26}$$

3. Calculate $c^{EB}(\alpha)$ as
$$c^{EB}(\alpha)=\text{conditional } (1-\alpha)\text{-quantile of } W^{EB} \text{ given } X_1^n. \tag{27}$$

We will call $c^{MB}(\alpha)$ and $c^{EB}(\alpha)$ the (one-step) Multiplier Bootstrap (MB) and Empirical Bootstrap (EB) critical values with size α. In practice conditional quantiles of W^{MB} or W^{EB} 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
$$\mathbb{E}_n\left[\frac{(X_{ij}-\widehat{\mu}_j)}{\widehat{\sigma}_j}\frac{(X_{ik}-\widehat{\mu}_k)}{\widehat{\sigma}_k}\right],\ 1\le j,k\le p, \tag{28}$$

---

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

[14] 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).

which should be close to the covariance matrix of the vector Y. Indeed, by Theorem 2 in Chernozhukov, Chetverikov, and Kato (2013b), the primary factor for the bound on the Kolmogorov distance between the conditional distribution of W and the distribution of $\max_{1\le j\le p} Y_j$ is

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

which we show to be small under suitable conditions even when $p \gg n$.

In turn, the empirical bootstrap is expected to work well since conditional on the data X_1^n, the maximum of the random vector

$$\left(\frac{\sqrt{n}\mathbb{E}_n[X_{ij}^*-\widehat{\mu}_j]}{\widehat{\sigma}_j}\right)_{1\le j\le p}$$

can be well approximated in distibution by the maximum of a random vector with centered normal distribution with covariance matrix (28) even when $p \gg n$.

The following theorem formally establishes validity of the MB and EB critical values.

Theorem 4.3 (Validity of one-step MB and EB methods)**.** *Let $c^B(\alpha)$ stay either for $c^{MB}(\alpha)$ or $c^{EB}(\alpha)$. Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$(M_{n,3}^3 \vee M_{n,4}^2 \vee B_n)^2 \log^{7/2}(pn) \le C_1 n^{1/2-c_1}. \tag{29}$$

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

$$\mathrm{P}(T > c^B(\alpha)) \le \alpha + Cn^{-c}. \tag{30}$$

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

$$|\mathrm{P}(T > c^B(\alpha)) - \alpha| \le Cn^{-c}. \tag{31}$$

Moreover, all these bounds hold uniformly with respect to the common distribution of X_i for which (12) and (29) are verified.

Comment 4.2 (High dimension bootstrap CLT)**.** The result (31) can be understood as a high dimensional bootstrap CLT for maxima of *studentized* sample averages. It shows that such maxima can be approximated either by multiplier or empirical bootstrap methods even if maxima is taken over (very) many sample averages. Moreover, the distributional approximation holds with polynomially (in n) small error. This result complements a high dimensional bootstrap CLT for *non-studentized* sample averages derived in Chernozhukov, Chetverikov, and Kato (2013a) and Chernozhukov, Chetverikov, and Kato (2014b), and may be of interest in many other settings, well beyond the problem of testing many moment inequalities.

Comment 4.3 (Other bootstrap procedures)**.** There exist many different bootstrap procedures in the literature, each with its own advantages and disadvantages. In this paper, we focused on multiplier and empirical bootstraps, and we leave analysis of more general exchangeably weighted bootstraps, which include many existing bootstrap procedures as a special case (see, for example, Praestgaard and Wellner (1993)), in the high dimensional setting for future work.

4.2.2. *Two-step methods.* We now consider to combine bootstrap methods with inequality selection. To describe these procedures, let $0 < \beta_n < \alpha/2$ be some constant. As in the previous section, we allow β_n to depend on n. Let $c^{MB}(\beta_n)$ and $c^{EB}(\beta_n)$ be one-step MB and EB critical values with size β_n, respectively. Define the sets $\widehat{J}_{MB}$ and $\widehat{J}_{EB}$ by

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

where B stays either for MB or EB. Then the two-step MB and EB critical values, $c^{MB,2S}(\alpha)$ and $c^{EB,2S}(\alpha)$, are defined by the following procedures:

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

1. Generate independent standard normal random variables $\epsilon_1,\dots,\epsilon_n$ independent of the data X_1^n.
2. Construct the multiplier 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 $c^{MB,2S}(\alpha)$ as
$$c^{MB,2S}(\alpha) = \text{conditional } (1-\alpha+2\beta_n)\text{-quantile of } W_{\widehat{J}_{MB}} \text{ given } X_1^n. \quad (32)$$

Algorithm (Empirical bootstrap with inequality selection)**.**

1. Generate a bootstrap sample $X_1^*,\dots,X_n^*$ as i.i.d. draws from the empirical distribution of $X_1^n = \{X_1,\dots,X_n\}$.
2. Construct the empirical bootstrap test statistic
$$W_{\widehat{J}_{EB}} = \begin{cases} \max_{j\in\widehat{J}_{EB}} \frac{\sqrt{n}\mathbb{E}_n[X_{ij}^*-\widehat{\mu}_j]}{\widehat{\sigma}_j}, & \text{if } \widehat{J}_{EB} \text{ is not empty,} \\ 0 & \text{if } \widehat{J}_{EB} \text{ is empty.} \end{cases}$$
3. Calculate $c^{EB,2S}(\alpha)$ as
$$c^{EB,2S}(\alpha) = \text{conditional } (1-\alpha+2\beta_n)\text{-quantile of } W_{\widehat{J}_{EB}} \text{ given } X_1^n. \quad (33)$$

The following theorem establishes validity of the two-step MB and EB critical values.

Theorem 4.4 (Validity of two-step MB and EB methods)**.** *Let* $c^{B,2S}(\alpha)$ *stay either for* $c^{MB,2S}(\alpha)$ *or* $c^{EB,2S}(\alpha)$*. Suppose that the assumption of Theorem 4.3 is satisfied. Moreover, suppose that* $\sup_{n\geq 1}\beta_n < \alpha/2$ *and* $\log(1/\beta_n) \leq C_1 \log n$*. Then there exist positive constants* c, C *depending only on* c_1, C_1

such that under H_0, $\mathrm{P}(T > c^{B,2S}(\alpha)) \leq \alpha + Cn^{-c}$. If $\mu_j = 0$ for all $1 \leq j \leq p$, then $\mathrm{P}(T > c^{B,2S}(\alpha)) \geq \alpha - 3\beta_n - Cn^{-c}$, so that if in addition $\beta_n \leq C_1 n^{-c_1}$, then $|\mathrm{P}(T > c^{B,2S}(\alpha)) - \alpha| \leq Cn^{-c}$. Finally, all these bounds hold uniformly with respect to the common distribution of X_i for which (12) and (29) are verified.

Comment 4.4. 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 an inequality 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 (31), 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 (15) with the following probabilistic inequality: under H_0,

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

where

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

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

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

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

4.3. **Hybrid methods.** We have considered the one-step SN, MB, and EB methods and their two-step variants. In fact, we can also consider "hybrids" of these methods. For example, we can use the SN method for inequality selection, and apply the MB or EB method for the selected inequalities, which is a computationally more tractable alternative to the two-step MB and EB methods. For convenience of terminology, we will call it the Hybrid (HB) method. To formally define the method, let $0 < \beta_n < \alpha/2$ be some constants, and recall the set $\widehat{J}_{SN} \subset \{1, \dots, p\}$ defined in (21). Suppose we want to use the MB method on the second step. Then the hybrid MB critical value, $c^{MB,H}(\alpha)$ is defined by the following procedure:

Algorithm (Multiplier Bootstrap 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 $c^{MB,H}(\alpha)$ as

$$c^{MB,H}(\alpha) = \text{conditional } (1-\alpha+2\beta_n)\text{-quantile of } W_{\widehat{J}_{SN}} \text{ given } X_1^n. \quad (34)$$

A similar algorithm can be defined for the EB method on the second step, which leads to the hybrid EB critical value $c^{EB,H}(\alpha)$. The following theorem establishes validity of these critical values.

Theorem 4.5 (Validity of hybrid two-step methods)**.** *Let $c^{B,H}(\alpha)$ stay either for $c^{MB,H}(\alpha)$ or $c^{EB,H}(\alpha)$. Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that (29) 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 4.4 hold with $c^{B,MS}(\alpha)$ replaced by $c^{B,H}(\alpha)$.*

4.4. **Three-step method.** In empirical studies based on moment inequalities, one typically has inequalities of the form

$$\mathrm{E}[g_j(\xi,\theta)] \leq 0 \quad \text{for all } j = 1,\dots,p, \quad (35)$$

where ξ is a vector of random variables, $\theta \in \mathbb{R}$ a vector of parameters, and $g_1,\dots,g_p$ a set of (known) functions. In these studies, inequalities (1)-(2) arise when one tests the null hypothesis $\theta = \theta_0$ against the alternative $\theta \neq \theta_0$ on the i.i.d. data $\xi_1,\dots,\xi_n$ by setting $X_{ij} := g_j(\xi_i,\theta_0)$ and $\mu_j := \mathrm{E}[X_{1j}]$. So far in this section, we showed how to increase power of such tests by employing inequality selection procedures that allow the researcher to drop uninformative inequalities, that is inequalities j with $\mu_j < 0$ if μ_j is not too close to 0. In this subsection, we seek to combine these selection procedures with another selection procedure that is suitable for the model (35) and that can increase local power of the test of $\theta = \theta_0$ by dropping *weakly informative* inequalities, that is inequalities j with the function $\theta \mapsto \mathrm{E}[g_j(\xi,\theta)]$ being flat or nearly flat around $\theta = \theta_0$. When the tested value θ_0 is close to some θ satisfying (35), such inequalities can only provide a weak signal of violation of the hypothesis $\theta = \theta_0$ in the sense that they have $\mu_j \approx 0$. For brevity of the paper, we only consider weakly informative inequality selection based on the MB and EB methods and note that similar results can be obtained for the self-normalized method.

We start with preparing necessary notation. Let $\xi_1,\dots,\xi_n$ be a sample of observations from the distribution of ξ. Suppose that we are interested in testing the null hypothesis

$$H_0 : \mathrm{E}[g_j(\xi,\theta_0)] \leq 0 \quad \text{for all } j = 1,\dots,p,$$

against the alternative

$$H_1 : \mathrm{E}[g_j(\xi,\theta_0)] > 0 \quad \text{for some } j = 1,\dots,p,$$

where θ_0 is some value of the parameter θ. Suppose that θ is a vector in $\mathbb{R}^r$. Denote

$$m_j(\xi,\theta) := \frac{\partial g_j(\xi,\theta)}{\partial\theta}$$

so that $m_j(\xi,\theta) = (m_{j1}(\xi,\theta),\dots,m_{jr}(\xi,\theta))^T$. Further, let $X_{ij} := g_j(\xi_i,\theta_0)$, $\mu_j := \mathrm{E}[X_{1j}]$, $\sigma_j := (\mathrm{Var}(X_{1j}))^{1/2}$, $V_{ijl} := m_{jl}(\xi_i,\theta_0)$, $\mu^V_{jl} := \mathrm{E}[V_{1jl}]$, and $\sigma^V_{jl} := (\mathrm{Var}(V_{1jl}))^{1/2}$. We assume that

$$\mathrm{E}[X_{1j}^2] < \infty,\ \sigma_j > 0,\ j = 1,\dots,p, \tag{36}$$

$$\mathrm{E}[V_{1jl}^2] < \infty,\ \sigma^V_{jl} > 0,\ j = 1,\dots,p,\ l = 1,\dots,r. \tag{37}$$

In addition, let

$$\widehat{\mu}_j = \mathbb{E}_n[X_{ij}] \text{ and } \widehat{\sigma}_j = \left(\mathbb{E}_n[(X_{ij}-\widehat{\mu}_j)^2]\right)^{1/2}$$

be estimators of μ_j and σ_j, respectively, and let

$$\widehat{\mu}^V_{jl} = \mathbb{E}_n[V_{ijl}] \text{ and } \widehat{\sigma}^V_{jl} = \left(\mathbb{E}_n[(V_{ijl}-\widehat{\mu}^V_{jl})^2]\right)^{1/2}$$

be estimators of μ^V_{jl} and σ^V_{jl}, respectively.

Weakly informative inequality selection that we derive is based on the bootstrap methods similar to those described in Section 4:

Algorithm (Multiplier bootstrap for gradient statistic)**.**

1. Generate independent standard normal random variables $\epsilon_1,\dots,\epsilon_n$ independent of the data $\xi_1^n = \{\xi_1,\dots,\xi_n\}$.

2. Construct the multiplier bootstrap gradient statistic

$$W^V_{MB} = \max_{j,l} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(V_{ijl}-\widehat{\mu}^V_{jl})]}{\widehat{\sigma}^V_{jl}}. \tag{38}$$

3. For $\gamma \in (0,1)$, calculate $c^{MB,V}(\gamma)$ as

$$c^{MB,V}(\gamma) = \text{conditional } (1-\gamma)\text{-quantile of } W^V_{MB} \text{ given } \xi_1^n. \tag{39}$$

Algorithm (Empirical bootstrap for gradient statistic)**.**

1. Generate a bootstrap sample $V_1^*,\dots,V_n^*$ as i.i.d. draws from the empirical distribution of $V_1^n = \{V_1,\dots,V_n\}$.

2. Construct the empirical bootstrap gradient statistic

$$W^V_{EB} = \max_{j,l} \frac{\sqrt{n}\mathbb{E}_n[V^*_{ijl}-\widehat{\mu}^V_{jl}]}{\widehat{\sigma}^V_{jl}}. \tag{40}$$

3. For $\gamma \in (0,1)$, calculate $c^{EB,V}(\gamma)$ as

$$c^{EB,V}(\gamma) = \text{conditional } (1-\gamma)\text{-quantile of } W^V_{EB} \text{ given } \xi_1^n. \tag{41}$$

Let φ_n be a sequence of constants satisfying $c_2 \le \varphi_n \log n \le C_2$ for some strictly positive constants c_2 and C_2, and let β_n be a sequence of constants satisfying $0 < \beta_n < \alpha/2$ and $\log(1/(\beta_n - \varphi_n)) \leq C_2 \log n$ where α is the nominal level of the test. Define three estimated sets of inequalities:

$$\widehat{J}_B := \left\{ j \in \{1, \dots, p\} : \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j > -2c^B(\beta_n) \right\},$$
$$\widehat{J}'_B := \left\{ j \in \{1, \dots, p\} : \sqrt{n}\widehat{\mu}^V_{jl}/\widehat{\sigma}^V_{jl} > -c^{B,V}(\beta_n + \varphi_n) \text{ for all } l = 1, \dots, r \right\},$$
$$\widehat{J}''_B := \left\{ j \in \{1, \dots, p\} : \sqrt{n}\widehat{\mu}^V_{jl}/\widehat{\sigma}^V_{jl} > -3c^{B,V}(\beta_n - \varphi_n) \text{ for all } l = 1, \dots, r \right\},$$

where B stays either for MB or EB.

Importantly, the weakly informative inequality selection procedure that we derive requires that both the test statistic and the critical value depend on the estimated sets of inequalities. Let T^B and $c^{B,3S}(\alpha)$ denote the test statistic and the critical value for $B = MB$ or EB depending on which bootstrap procedure is used. If the set $\widehat{J}'_B$ is empty, set the test statistic $T^B = 0$ and the critical value $c^{B,3S}(\alpha) = 0$. Otherwise, define the test statistic

$$T^B = \max_{j \in \widehat{J}'_B} \frac{\sqrt{n}\widehat{\mu}_j}{\widehat{\sigma}_j},$$

and define the three-step MB/EB critical values, $c^{B,3S}(\alpha)$ for the test by the same bootstrap procedures as those for $c^{B,2S}(\alpha)$ with $\widehat{J}_B$ replaced by $\widehat{J}_B \cap \widehat{J}''_B$, and also $2\beta_n$ replaced by $4\beta_n$ (in (32) and (33)) (Algorithms "Multiplier bootstrap with inequality selection" and "Empirical bootstrap with inequality selection"). The test rejects H_0 if $T^B > c^{B,3S}(\alpha)$.

To state the main result of this section, we need the following additional notation. Let

$$Z^V_{ijl} := (V_{ijl} - \mu^V_{jl})/\sigma^V_{jl}.$$

Observe that $\mathrm{E}[Z^V_{ijl}] = 0$ and $\mathrm{E}[(Z^V_{ijl})^2] = 1$. Let

$$M^V_{n,k} := \max_{j,l} \left(\mathrm{E}[|Z^V_{1jl}|^k] \right)^{1/k}, \ k = 3, 4, \ B^V_n := \left(\mathrm{E}\Big[\max_{j,l} (Z^V_{1jl})^4 \Big] \right)^{1/4}.$$

We have the following theorem:

Theorem 4.6 (Validity of three-step MB and EB methods)**.** *Let T^B and $c^{B,3S}(\alpha)$ stay either for T^{MB} and $c^{MB,3S}(\alpha)$ or for T^{EB} and $c^{EB,3S}(\alpha)$. Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$\left(M^3_{n,3} \vee M^2_{n,4} \vee B_n \vee (M^V_{n,3})^3 \vee (M^V_{n,4})^2 \vee B^V_n \right)^2 \log^{7/2}(prn) \leq C_1 n^{1/2 - c_1}. \tag{42}$$

Moreover, suppose that $\sup_{n \geq 1} \beta_n < \alpha/2$, $\log(1/(\beta_n - \varphi_n)) \leq C_2 \log n$, and $c_2 \leq \varphi_n \log n \leq C_2$. Then there exist positive constants c, C depending only on c_1, C_1, c_2, and C_2 such that under H_0, $\mathrm{P}(T^B > c^{B,3S}(\alpha)) \leq \alpha + Cn^{-c}$. In addition, the bound holds uniformly with respect to the common distribution of ξ_i for which (36), (37), and (42) are verified.

Comment 4.5 (On the choice of φ_n)**.** Inspecting the proof of the theorem shows that the result of the theorem remains valid if we replace condition

$c_2 \leq \varphi_n \log n \leq C_2$ for some c_2, C_2 by weaker conditions $\varphi_n \to 0$ and $\varphi_n \geq Cn^{-c}$ for some constants c, C that can be chosen to depend only on c_1, C_1. In practice, however, it is difficult to track the dependence of c, C on c_1, C_1. Therefore, in the main text we state the result with the condition $c_2 \leq \varphi_n \log n \leq C_2$.

5. 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 begin with deriving an upper bound on the power any procedure may have in testing (1) against (2).

Lemma 5.1 (Upper bounds on power). *Let $X_1, \dots, X_n \sim N(\mu, \Sigma)$ be 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 $H_1 : \max_{1 \leq j \leq p} (\mu_j / \sigma_j) \geq r$ with $r > 0$ 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*

$$\inf_{\max_{1 \leq j \leq p}(\mu_j/\sigma_j) \geq r} \mathrm{E}_\mu[\phi_n(X_1, \dots, X_n)] \leq \alpha + \mathrm{E}\left[|p^{-1} \textstyle\sum_{j=1}^p e^{\sqrt{n} r \xi_j - n r^2/2} - 1| \right], \tag{43}$$

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

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

where $\underline{r}_n = (1 - \epsilon_n)\sqrt{2(\log p_n)/n}$, 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 (12) and consider the test statistic T defined in (13). 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 5.2 (Lower bounds on power). *In the setting described above, for every $\underline{\epsilon} \geq 0$, there exist $\epsilon > 0$ and $\delta \in (0,1)$ such that whenever*

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

we have

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

From this lemma, we have the following corollary:

Corollary 5.1 (Asymptotic minimax optimality). *Let $\widehat{c}(\alpha)$ be any one of $c^{SN}(\alpha)$, $c^{SN,2S}(\alpha)$, $c^{MB}(\alpha)$, $c^{MB,2S}(\alpha)$, $c^{EB}(\alpha)$, $c^{EB,2S}(\alpha)$, $c^{MB,H}(\alpha)$, or $c^{EB,H}(\alpha)$ where we assume $\sup_{n\geq 1}\beta_n \leq \alpha/3$ whenever inequality selection is used. Suppose there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$M_{n,4}^2 \log^{1/2} p \leq C_1 n^{1/2-c_1}, \text{ and } \log^{3/2} p \leq C_1 n. \tag{44}$$

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)

$$\inf_{\max_{1\leq j\leq p}(\mu_j/\sigma_j)\geq \overline{r}_n} \mathrm{P}_\mu(T > \widehat{c}(\alpha)) \geq 1 - o(1), \tag{45}$$

where $\overline{r}_n = (1+\epsilon_n)\sqrt{2(\log p_n)/n}$ and P_μ is the probability under μ. Moreover, the above asymptotic result (45) holds uniformly with respect to the sequence of common distributions of X_i for which (12) and (44) are verified with given c_1, C_1.

Comparing the bounds in Lemma 5.1 and Corollary 5.1, we see that the tests developed in this paper are asymptotically minimax optimal when $p = p_n \to \infty$ as $n \to \infty$ under mild regularity conditions.

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

In this section, 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 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

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

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 $n \geq 1$,

$$\inf_{P \in \mathcal{P}_n} \inf_{\theta \in \Theta_0(P)} \mathrm{P}_P(\theta \in \mathcal{C}_n(\alpha)) \geq 1 - \alpha - Cn^{-c}, \tag{46}$$

while allowing for $p > n$ (indeed we allow p to be much larger than n), where $0 < \alpha < 1/2$ and $\mathcal{P}_n$ is a suitable sequence of classes of distributions on $(S, \mathcal{S})$. We call confidence regions $\mathcal{C}_n(\alpha)$ for which (46) 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 (2014a).

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

$$\begin{gathered}\Theta_0(P) \neq \emptyset, \text{ and } \mathrm{E}_P[g_j^2(\xi_1, \theta)] < \infty, \ \sigma_j^2(\theta, P) := \mathrm{Var}_P(g_j(\xi_1, \theta)) > 0, \\ \text{for all } j = 1, \dots, p, \text{ and all } \theta \in \Theta_0(P).\end{gathered} \tag{47}$$

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 of testing

$$H_\theta : \mu_j(\theta, P) \leq 0, \text{ for all } j = 1, \dots, p,$$

against the alternative

$$H_\theta' : \mu_j(\theta, P) > 0, \text{ for some } j = 1, \dots, p,$$

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

$$\mathcal{C}_n(\alpha) = \{\theta \in \Theta : T(\theta) \leq c(\alpha, \theta)\}, \tag{48}$$

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 (18), and let $c^{SN,2S}(\alpha, \theta)$, $c^{MB}(\alpha, \theta)$, $c^{MB,2S}(\alpha, \theta)$, $c^{EB}(\alpha, \theta)$, $c^{EB,2S}(\alpha, \theta)$, $c^{MB,H}(\alpha, \theta)$, and $c^{EB,H}(\alpha, \theta)$ be the two-step SN, one-step MB, two-step MB, one-step EB, two-step EB, MB hybrid, and EB hybrid critical values defined in Section 4 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 (48) with $c(\alpha, \theta) = c^{SN}(\alpha)$; define

$$\mathcal{C}_n^{SN,2S}(\alpha), \mathcal{C}_n^{MB}(\alpha), \mathcal{C}_n^{MB,2S}(\alpha), \mathcal{C}_n^{EB}(\alpha), \mathcal{C}_n^{EB,2S}(\alpha), \mathcal{C}_n^{MB,H}(\alpha), \mathcal{C}_n^{EB,H}(\alpha)$$

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. The following theorem is the main result of this section.

Theorem 6.1. *Let $\mathcal{P}_n^{SN}$ be the class of distributions P on $(S, \mathcal{S})$ for which (47) and (20) are verified with $M_{n,3}$ replaced by $M_{n,3}(\theta, P)$ for all $\theta \in \Theta_0(P)$; let $\mathcal{P}_n^{SN,2S}$ be the class of distributions P on $(S, \mathcal{S})$ for which (47) and (23) 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 $\mathcal{P}_n^B$ be the class of distributions P on $(S, \mathcal{S})$ for which (47) and (29) are verified with $M_{n,k}, B_n$ replaced by (respectively) $M_{n,k}(\theta, P), B_n(\theta, P)$ for all $\theta \in \Theta_0(P)$.*[15] *Moreover, suppose that $\sup_{n\geq 1} \beta_n \leq \alpha/3$ and $\log(1/\beta_n) \leq C_1 \log n$ whenever inequality selection is used. Then there exist positive constants c, C depending only on α, c_1, C_1 such that*

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

where $(\mathcal{P}_n, \mathcal{C}_n)$ is one of the pairs $(\mathcal{P}_n^{SN}, \mathcal{C}_n^{SN})$, $(\mathcal{P}_n^{SN,2S}, \mathcal{C}_n^{SN,2S})$, $(\mathcal{P}_n^B, \mathcal{C}_n^{MB})$, $(\mathcal{P}_n^B, \mathcal{C}_n^{MB,2S})$, $(\mathcal{P}_n^B, \mathcal{C}_n^{EB})$, $(\mathcal{P}_n^B, \mathcal{C}_n^{EB,2S})$, $(\mathcal{P}_n^B, \mathcal{C}_n^{MB,H})$ or $(\mathcal{P}_n^B, \mathcal{C}_n^{EB,H})$.

7. Extensions

7.1. **Dependent data.**

In this section we consider the case where $X_1, \dots, X_n$ are dependent. To avoid technical complication, we focus here on the non-Studentized version of T:

$$\check{T} = \max_{1\leq j\leq p} \sqrt{n}\widehat{\mu}_j.$$

We consider a version of the multiplier bootstrap, namely the block multiplier bootstrap, to calculate critical values for $\check{T}$, where a certain blocking technique is used to account for dependency among $X_1, \dots, X_n$.[16]

Let $X_1, \dots, X_n$ be possibly dependent random vectors in $\mathbb{R}^p$ with identical distribution (that is, $X_i \stackrel{d}{=} X_1$, for all $i = 1, \dots, n$), defined on the probability space $(\Omega, \mathcal{A}, \mathrm{P})$. We follow the basic notation introduced in Section 3. For the sake of simplicity, assume that there exists a constant $D_n \geq 1$ such that $|X_{ij} - \mu_j| \leq D_n$ a.s. for $1 \leq i \leq n, 1 \leq j \leq p$.

For any integer $1 \leq q \leq n$, define

$$\overline{\sigma}^2(q) := \max_{1\leq j\leq p} \max_I \mathrm{Var}\left(q^{-1/2} \sum_{i\in I} X_{ij}\right),$$

$$\underline{\sigma}^2(q) := \min_{1\leq j\leq p} \min_I \mathrm{Var}\left(q^{-1/2} \sum_{i\in I} X_{ij}\right),$$

[15] For example, $\mathcal{P}_n^{SN} = \{P : \text{(47) is verified, and } M_{n,3}^3(\theta, P) \log^{3/2}(p/\alpha) \leq C_1 n^{1/2-c_1}, \forall\theta \in \Theta_0(P)\}$.

[16] We refer to Lahiri (2003) as a general reference on resampling methods for dependent data.

where $\max_I$ and $\min_I$ are taken over all $I \subset \{1, \dots, n\}$ of the form $I = \{i+1, \dots, i+q\}$. For any sub σ-fields $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$, define

$$\beta(\mathcal{A}_1, \mathcal{A}_2) := \frac{1}{2} \sup \Bigg\{ \sum_{i=1}^{I} \sum_{j=1}^{J} |\mathrm{P}(A_i \cap B_j) - \mathrm{P}(A_i)\mathrm{P}(B_j)| : \{A_i\} \text{ is any partition of } \Omega \text{ in } \mathcal{A}_1, \{B_j\} \text{ is any partition of } \Omega \text{ in } \mathcal{A}_2 \Bigg\}.$$

Define the kth β-mixing coefficient for $X_1^n = \{X_1, \dots, X_n\}$ by

$$b_k = b_k(X_1^n) = \max_{1 \le l \le n-k} \beta(\sigma(X_1, \dots, X_l), \sigma(X_{l+k}, \dots, X_n)),\ 1 \le k \le n-1,$$

where $\sigma(X_i, i \in I)$ with $I \subset \{1, \dots, n\}$ is the σ-field generated by $X_i, i \in I$.[17]

We employ Bernstein's "small-block and large-block" technique and decompose the sequence $\{1, \dots, n\}$ into "large" and "small" blocks. Let $q > r$ be positive integers with $q + r \le n/2$ (q, r depend on n: $q = q_n, r = r_n$, and asymptotically we require $q_n \to \infty, q_n = o(n), r_n \to \infty$, and $r_n = o(q_n)$), and let $I_1 = \{1, \dots, q\}, J_1 = \{q+1, \dots, q+r\}, \dots, I_m = \{(m-1)(q+r)+1, \dots, (m-1)(q+r)+q\}, J_m = \{(m-1)(q+r)+q+1, \dots, m(q+r)\}, J_{m+1} = \{m(q+r), \dots, n\}$, where $m = m_n = [n/(q+r)]$ (the integer part of $n/(q+r)$).

Then the block multiplier bootstrap is described as follows: generate independent standard normal random variables $\epsilon_1, \dots, \epsilon_m$, independent of X_1^n. Let

$$\check{W} = \max_{1 \le j \le p} \frac{1}{\sqrt{mq}} \sum_{l=1}^{m} \epsilon_l \sum_{i \in I_l} (X_{ij} - \widehat{\mu}_j),$$

and consider

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

which we call the BMB (Block Multiplier Bootstrap) critical value.

Theorem 7.1 (Validity of BMB method)**.** *Work under the setting described above. Suppose that there exist constants $0 < c_1 \le C_1$ and $0 < c_2 < 1/4$ such that $c_1 \le \underline{\sigma}^2(q) \le \overline{\sigma}^2(r) \vee \overline{\sigma}^2(q) \le C_1, \max\{mb_r, (r/q)\log^2 p\} \le C_1 n^{-c_2}$, and $qD_n \log^{5/2}(pn) \le C_1 n^{1/2-c_2}$. Then there exist positive constants c, C depending only on c_1, c_2, C_1 such that under H_0, $\mathrm{P}(\check{T} > \widehat{c}^{BMB}(\alpha)) \le \alpha + Cn^{-c}$. If $\mu_j = 0$ for all $1 \le j \le p$, then $|\mathrm{P}(\check{T} > \widehat{c}^{BMB}(\alpha)) - \alpha| \le Cn^{-c}$.*

Comment 7.1 (Connection to tapered block bootstrap)**.** The BMB method can be considered as a variant of the tapered block bootstrap (see Paparoditis and Politis, 2001, 2002; Andrews, 2004) applied to non-overlapping blocks with a rectangular tapering function. The difference is that in the original tapered block bootstrap the multipliers are multinomially distributed, while in the BMB the multipliers are independent standard normal.

[17]We refer to Fan and Yao (2003), Section 2.6, as a general reference on mixing.

7.2. **Approximate moment inequalities.** As shown in a dynamic model of imperfect competition example in Section 2.3, in some applications, random vectors $X_1, \dots, X_n$ satisfying inequalities (1) with $\mu_j = \mathrm{E}[X_{1j}]$ are not observed. Instead, the data consist of random vectors $\widehat{X}_1, \dots, \widehat{X}_n$ that approximate vectors $X_1, \dots, X_n$. In that example, the approximation error arises from the need to linearize original inequalities. Another possibility leading to nontrivial approximation error is that where the data contain estimated parameters. In this section, we derive a set of conditions that suffice for the same results as those obtained in Section 4 when we use the data $\widehat{X}_1, \dots, \widehat{X}_n$ as if we were using exact vectors $X_1, \dots, X_n$. For brevity, we only consider two-step MB/EB methods.

We use the following notation. Let $\widehat{\mu}_{j,0} := \mathbb{E}_n[X_{ij}]$ and $\widehat{\sigma}^2_{j,0} := \mathbb{E}_n[(X_{ij} - \widehat{\mu}_{j,0})^2]$ denote (infeasible) estimators of $\mu_j = \mathrm{E}[X_{1j}]$ and $\sigma_j^2 = \mathrm{Var}(X_{1j})$. In addition, assume that we have estimates $\widehat{\mu}_j$ that appropriately approximate $\widehat{\mu}_{j,0}$ for $j = 1, \dots, p$. In the context of Section 2.3, for example, these estimates would take the form $\widehat{V}_j(s, \sigma_j', \sigma_{-j}(\widehat{\alpha}_n), \theta) - \widehat{V}_j(s, \sigma_j(\widehat{\alpha}_n), \sigma_{-j}(\widehat{\alpha}_n), \theta)$. Moreover, let $\widehat{\sigma}_j^2 := \mathbb{E}_n[(X_{ij} - \widehat{\mu}_j)^2]$ be a (feasible) estimator of σ_j^2.

Define the test statistic T by (13); that is, $T = \max_{1 \le j \le p} \sqrt{n}\widehat{\mu}_j / \widehat{\sigma}_j$. Define the critical value $c^{B,2S}(\alpha)$ for $B = MB$ or EB by the same algorithms as those used in Section 4 with X_{ij} replaced by $\widehat{X}_{ij}$ for all i and j (and using $\widehat{\mu}_j$ and $\widehat{\sigma}_j^2$ as defined in this section). We have the following theorem:

Theorem 7.2 (Validity of two-step MB/EB methods for approximate inequalities)**.** *Let $c^{B,2S}(\alpha)$ stay either for $c^{MB,2S}(\alpha)$ or $c^{EB,2S}(\alpha)$. Suppose that the assumption of Theorem 4.3 is satisfied. Moreover, suppose that $\sup_{n \ge 1} \beta_n \le \alpha/2$ and $\log(1/\beta_n) \le C_1 \log n$. In addition, suppose that there exists a sequence ζ_{n1} satisfying $\zeta_{n1} \log p \le C_1 n^{-c_1}$ and such that $\mathrm{P}(\max_{1 \le j \le p} \sqrt{n}|\widehat{\mu}_j - \widehat{\mu}_{j,0}| > \zeta_{n1}) \le C_1 n^{-c_1}$ and $\mathrm{P}(\max_{1 \le j \le p} (\mathbb{E}_n[(\widehat{X}_{ij} - X_{ij})^2])^{1/2} > \zeta_{n1}) \le C_1 n^{-c_1}$. Moreover, if the EB method is used, suppose that $\mathrm{P}(\sqrt{\log p} \max_{i,j} |\widehat{X}_{ij} - X_{ij}| > \sqrt{n}\zeta_{n,1}) \le C_1 n^{-c_1}$. Finally, assume that $\sigma_j \ge c_1$ for all $j = 1, \dots, p$. Then all the conclusions of Theorem 4.4 hold with T, $c^{MB,2S}(\alpha)$, and $c^{EB,2S}(\alpha)$ defined in this section.*

8. Monte Carlo Experiments

In this section, we provide results of a small Monte Carlo simulation study. Our simulations demonstrate that the methods developed in this paper, on the one hand, have good size control and, on the other hand, have good power properties even though we use experimental setups with large number of moment inequalities.

Throughout all the experiments, we consider i.i.d. samples of size $n = 400$. Depending on the experiment, the number of moment inequalities is $p = 200$, 500, or 1000. Thus, we consider models where the number of moment inequalities p is comparable, larger, or substantially larger than the sample size n.

We consider eight different experimental designs. Designs 1-4 satisfy the null hypothesis H_0 but Designs 5-8 do not. Designs 1, 2, 5, and 6 are based on equicorrelated data where $X_i = \mu + A^T \varepsilon_i$ with $\varepsilon_i := (\varepsilon_{i1}, \dots, \varepsilon_{ip})^T$ being a vector consisting of i.i.d. random variables, $\Sigma := \mathrm{Var}(X_i) = A^T A$ being a matrix defined by $\Sigma_{jj} = 1$ and $\Sigma_{jk} = \rho$ for $1 \leq j, k \leq p$ and $j \neq k$, and $\mu = (\mu_1, \dots, \mu_p)^T$ being a non-stochastic vector representing the mean of X_i. Designs 3, 4, 7, and 8 are based on autocorrelated data where $X_i = \mu + A^T \varepsilon_i$ with ε_i and μ as above and $\Sigma := \mathrm{Var}(X_i) = A^T A$ being a matrix defined by $\Sigma_{jk} = \rho^{|j-k|}$ for $1 \leq j, k \leq p$. Depending on the experiment, we set $\rho = 0$, 0.5, or 0.9, and we consider ε_{ij} either having the normalized Student's t distribution with 4 degrees of freedom (we normalized the distribution to have variance one by dividing it by $\sqrt{2}$; in the tables below, this distribution is denoted as $t(4)/\sqrt{2}$) or having the uniform distribution on the interval $(-\sqrt{3}, \sqrt{3})$ (in the tables below, this distribution is denoted as $\sqrt{3}U(-1, 1)$). Thus, in all cases $\mathrm{Var}(X_{ij}) = 1$ for all $1 \leq j \leq p$.

In Designs 1 and 3, we set $\mu_j = 0$ for all $1 \leq j \leq p$. In Designs 2 and 4, we set $\mu_j = 0$ for $1 \leq j \leq \gamma p$ and $\mu_j = -0.8$ for $\gamma p + 1 \leq j \leq p$. Thus, Designs 1-4 satisfy H_0. In Designs 5 and 7, we set $\mu_j = 0.05$ for all $1 \leq j \leq p$. In Designs 6 and 8, we set $\mu_j = 0.05$ for $1 \leq j \leq \gamma p$ and $\mu_j = -0.75$ for $\gamma p + 1 \leq j \leq p$. Thus, Designs 5-8 do not satisfy H_0. In all experiments, we set $\gamma = 0.1$, so that in Designs 6 and 8, only 10% of inequalities violate H_0.

We consider self-normalized (SN), multiplier bootstrap (MB), and empirical bootstrap (EB) critical values, with and without inequality selection. In all experiments, we set the nominal level of the test $\alpha = 5\%$ and for tests with inequality selection, we set $\beta = 0.1\%$. We present results based on 1000 simulations for each design and we use $B = 1000$ bootstrap samples for each bootstrap procedure.

Results on the probabilities of rejecting H_0 in all the experiments are presented in Tables 1-4. The first observation that can be taken from these tables is that MB and EB methods give similar results. Therefore, in what follows we discuss and compare SN and bootstrap (MB and EB) critical values.

Tables 1 and 2 give results for Designs 1-4, where H_0 holds, and demonstrate that all of our tests have good size control. The largest over-rejection occurs in Design 3 with autocorrelated data, uniform ε_{ij}'s, $p = 1000$, and $\rho = 0.5$ where the MB and EB tests without inequality selection reject H_0 with probability 7.6% and 8.0% against the nominal level $\alpha = 5\%$, respectively. As expected, the self-normalized test tend to under-reject H_0 but the bootstrap tests take the correlation structure of the data into account, and have rejection probability close to nominal level $\alpha = 5\%$ in Designs 1 and 3, where inequalities hold as equalities. The most striking difference between the SN and bootstrap tests in this dimension perhaps can be seen in Design 1 with equicorrelated data, uniform ε_{ij}'s, $p = 1000$, and $\rho = 0.9$ where the MB and EB tests with selection reject H_0 with probability 5.3% and 5.4%, which is very close to the nominal level $\alpha = 5\%$, but the SN test with selection only rejects H_0 with probability 0.0%. Observe also that when the correlation in the data is not too large, the SN tests also have size rather close to the nominal level; see results for Design 3 with autocorrelated data and $\rho = 0$ or 0.5.

Tables 3 and 4 give results for designs 5-8, where H_0 does not hold, and demonstrate power properties of our tests. To understand the results, note first that there are no other tests in the literature that could be used in our experiments since we have (very) large number of moment inequalities p. As a benchmark, we thus compare the rejection probabilities with those we would obtain should we had only one moment inequality to test. For these purposes, we have chosen the mean-value $\mu_j = 0.05$ for the moment inequalities j violating H_0 so that $0.05 = (\text{Var}(X_{ij})/n)^{1/2}$ since $n = 400$ and $\text{Var}(X_{ij}) = 1$. This is a typical value of the moment that one needs to obtain a non-trivial power when only one moment inequality is tested. Indeed, the standard one-sided t-test would reject H_0 with probability approximately 25% in this case. Tables 3 and 4 thus show that our tests have good power. In particular, the rejection probability for the bootstrap tests with selection never falls below 25% and in some cases approaches 100%; see results for Designs 5 and 7 with $p = 1000$ and $\rho = 0$. The SN tests have rejection probabilities close to those for the bootstrap tests when $\rho = 0$ or even when $\rho = 0.5$ for Designs 7 and 8 with autocorrelated data. Our tests have good power even though we have a large number of moment inequalities p and in some cases (Designs 6 and 8) only a small fraction (10%) of these inequalities violate H_0. Further, the bootstrap tests substantially improve upon the SN test in cases with large correlation in the data; see, for example, results for Design 5 with equicorrelated data, ε_{ij} having Student's t-distribution, $p = 1000$ and $\rho = 0.5$, where the SN tests reject H_0 with probability 19% and the MB and EB tests reject H_0 with probability 37-39%. Finally, selection procedures yield important power improvements; see, for example, results for Design 8 with autocorrelated data, ε_{ij} having Student's t-distribution, $p = 1000$ and $\rho = 0.5$, where the MB test without selection rejects H_0 with probability 17% and the MB test with selection rejects H_0 with probability 66%.

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* $0 \le x \le n^{\frac{\nu}{2(2+\nu)}}/D_{n,\nu}$,

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

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* $t>0$,

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

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,

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

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). $\square$

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 $s \ge 1$. Then

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

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

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

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 $1 \le j \le p$. (i) For $\alpha \in (0,1)$, let $c_0(\alpha)$ denote the $(1-\alpha)$-quantile of the distribution of $\max_{1\le j\le p} Y_j$. Then $c_0(\alpha) \le \sqrt{2\log p} + \sqrt{2\log(1/\alpha)}$. (ii) For every $t \in \mathbb{R}$ and $\epsilon > 0$, $\mathrm{P}(|\max_{1\le j\le p} Y_j - t| \le \epsilon) \le 4\epsilon(\sqrt{2\log p} + 1)$.*

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

$$\mathrm{E}\left[\max_{1\le j\le p} Y_j\right] \le \sqrt{2\log p}. \tag{49}$$

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

$$\mathrm{P}\left(\max_{1\le j\le p} Y_j \ge \mathrm{E}\Big[\max_{1\le j\le p} Y_j\Big] + r\right) \le e^{-r^2/2},$$

by which we have

$$c_0(\alpha) \le \mathrm{E}\left[\max_{1\le j\le p} Y_j\right] + \sqrt{2\log(1/\alpha)}. \tag{50}$$

Combining (50) and (49) leads to the desired result. $\square$

A.2. **Proof of Theorem 4.1.** The first assertion follows from inequality (17) and Lemma A.1 with $\nu = 1$. To prove the second assertion, we first note the well known fact that $1 - \Phi(t) \le e^{-t^2/2}$ for $t > 0$, by which we have $\Phi^{-1}(1-\alpha/p) \le \sqrt{2\log(p/\alpha)}$.[18] Hence if $M_{n,3}^3 \log^{3/2}(p/\alpha) \le C_1 n^{1/2-c_1}$, it

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

is straightforward to verify that the right side on (19) is bounded by Cn^{-c_1} for some constant C depending only on c_1, C_1. □

A.3. **Proof of Theorem 4.2.** We first prove the following technical lemma. Recall that $B_n = (\mathrm{E}[\max_{1\leq j\leq p} Z_{1j}^4])^{1/4}$.

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

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

where K, K' *are universal constants.*

Proof. Here $K_1, K_2, \dots$ denote universal positive constants. Note that for $a > 0$, $|\sqrt{a} - 1| = |a-1|/(\sqrt{a}+1) \leq |a-1|$, so that for $r > 0$,

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

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

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

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

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

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

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

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

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

$$\mathrm{P}\left(\max_{1\leq j\leq p} |\mathbb{E}_n[Z_{ij}^2] - 1| > K_4 n^{-(1-c)/2} B_n^2 (\log p)\right) \leq K_3 n^{-c}. \tag{51}$$

Similarly, using Lemma A.3, we have

$$\mathrm{E}\left[\max_{1\leq j\leq p} |\mathbb{E}_n[Z_{ij}]|\right] \leq K_5(n^{-1/2}\sqrt{\log p} + n^{-3/4}B_n \log p), \tag{52}$$

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

$$\begin{aligned}\mathrm{P}&\left(\max_{1\le j\le p}|\mathbb{E}_n[Z_{ij}]| > 2K_5(n^{-1/2}\sqrt{\log p}+n^{-3/4}B_n\log p)+t\right)\\&\le e^{-nt^2/3}+K_6t^{-4}n^{-3}B_n^4.\end{aligned}$$

Taking $t = n^{-1/4}B_n$, the right-hand side becomes $e^{-n^{1/2}B_n/3} + K_6n^{-2} \le K_7n^{-2}$. Hence we have

$$\mathrm{P}\left(\max_{1\le j\le p}|\mathbb{E}_n[Z_{ij}]| > K_8(n^{-1/4}B_n\sqrt{\log p}+n^{-3/4}B_n\log p)\right)\le K_7n^{-2}. \quad (53)$$

Combining (51) and (53) leads to the desired result. □

Proof of Theorem 4.2. Here c, C denote generic positive constants depending only on α, c_1, C_1; their values may change from place to place. Define

$$J_1 = \{j\in\{1,\dots,p\} : \sqrt{n}\mu_j/\sigma_j > -c^{SN}(\beta_n)\},\ J_1^c = \{1,\dots,p\}\backslash J_1. \quad (54)$$

For $k \ge 1$, let

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

Note that $c^{SN,2S}(\alpha) = c^{SN,2S}(\alpha,\widehat{k})$ 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

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

so that it is enough to prove that

$$\mathrm{P}\left(\max_{1\le j\le p}\sqrt{n}(\widehat{\mu}_j-\mu_j)/\sigma_j > c^{SN}(\beta_n)\right)\le \beta_n + Cn^{-c}. \quad (55)$$

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-hand side of (55) is bounded by

$$\mathrm{P}\left(\max_{1\le j\le p}\sqrt{n}(\widehat{\mu}_j-\mu_j)/\widehat{\sigma}_j > (1-r)c^{SN}(\beta_n)\right) \quad (56)$$

$$+\,\mathrm{P}\left(\max_{1\le j\le p}|(\sigma_j/\widehat{\sigma}_j)-1| > r\right), \quad (57)$$

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 probability in (57) is bounded by Cn^{-c}.

Consider the probability in (56). It is not difficult to see that

$$\begin{aligned}
&\mathrm{P}\left(\max_{1\le j\le p}\sqrt{n}(\widehat{\mu}_j-\mu_j)/\widehat{\sigma}_j>(1-r)c^{SN}(\beta_n)\right)\\
&\quad\le \mathrm{P}\left(\max_{1\le j\le p}U_j>(1-r)\Phi^{-1}(1-\beta_n/p)\right)\\
&\quad\le \sum_{j=1}^{p}\mathrm{P}\left(U_j>(1-r)\Phi^{-1}(1-\beta_n/p)\right). \qquad (58)
\end{aligned}$$

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 sum in (58) is bounded by

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

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 by Taylor's expansion for some $r'\in[0,r]$,

$$p\bar{\Phi}\left((1-r)\xi\right)=\beta_n+rp\xi\phi\left((1-r')\xi\right)\le\beta_n+rp\xi\phi\left((1-r)\xi\right).$$

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 $\phi(\xi)\le 2\xi(1-\Phi(\xi))=2\xi\beta_n/p$.[19] 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 probability in (56) 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

$$\mathrm{P}(\widehat{J}_{SN}\not\supset J_1)\le\mathrm{P}\left(\max_{1\le j\le p}\left[\sqrt{n}(\mu_j-\widehat{\mu}_j)-(2\widehat{\sigma}_j-\sigma_j)c^{SN}(\beta_n)\right]>0\right). \qquad (59)$$

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

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

[19]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-hand side on (59) 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 4.2. Consider first the case where $J_1=\emptyset$. Then by Step 1, with probability larger than $1-\beta_n-Cn^{-c}$, $T\le 0$, so that

$$\mathrm{P}(T>c^{SN,2S}(\alpha))\le\beta_n+Cn^{-c}\le\alpha+Cn^{-c}.$$

Suppose now that $|J_1|\ge 1$. Observe that

$$\{T>c^{SN,2S}(\alpha)\}\cap\left\{\max_{j\in J_1^c}\widehat{\mu}_j\le 0\right\}\subset\left\{\max_{j\in J_1}\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j>c^{SN,2S}(\alpha)\right\}.$$

Moreover, as $c^{SN,2S}(\alpha,k)$ is non-decreasing in k,

$$\left\{\max_{j\in J_1}\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j>c^{SN,2S}(\alpha)\right\}\cap\{\widehat{J}_{SN}\supset J_1\}$$
$$\subset\left\{\max_{j\in J_1}\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j>c^{SN,2S}(\alpha,|J_1|)\right\}.$$

Therefore, by Steps 1 and 2, we have

$$\begin{aligned}&\mathrm{P}(T>c^{SN,2S}(\alpha))\\&\quad\le\mathrm{P}\left(\max_{j\in J_1}\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j>c^{SN,2S}(\alpha,|J_1|)\right)+2\beta_n+Cn^{-c}\\&\quad\le\mathrm{P}\left(\max_{j\in J_1}\sqrt{n}(\widehat{\mu}_j-\mu_j)/\widehat{\sigma}_j>c^{SN,2S}(\alpha,|J_1|)\right)+2\beta_n+Cn^{-c}.\end{aligned}\tag{60}$$

By Theorem 4.1, we see that

$$\mathrm{P}\left(\max_{j\in J_1}\sqrt{n}(\widehat{\mu}_j-\mu_j)/\widehat{\sigma}_j>c^{SN,2S}(\alpha,|J_1|)\right)\le\alpha-2\beta_n+Cn^{-c}.\tag{61}$$

Combining (60) and (61) completes the proof of the theorem. □

A.4. **Proof of Theorem 4.3.** Here c,C denote generic positive constants depending only on c_1,C_1; their values may change from place to place. Let W stay for W^{MB} or W^{EB}, depending on which bootstrap procedure is used. Define

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

In addition, define

$$\bar{W}^{MB} := \max_{1\le j\le p} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_j)]}{\sigma_j},\ \bar{W}^{EB} := \max_{1\le j\le p} \frac{\sqrt{n}\mathbb{E}_n[(X^*_{ij}-\widehat{\mu}_j)]}{\sigma_j},$$

and let $\bar{W}$ stay for $\bar{W}^{MB}$ or $\bar{W}^{EB}$ depending on which bootstrap procedure is used. Further, let

$$(Y_1,\dots,Y_p)^T \sim N(0,\mathrm{E}[Z_1Z_1^T])$$

and for $\gamma\in(0,1)$, denote by $c_0(\gamma)$ the $(1-\gamma)$-quantile of the distribution of $\max_{1\le j\le p} Y_j$. Finally, define

$$\rho_n := \sup_{t\in\mathbb{R}} \left| \mathrm{P}(T_0\le t \mid X_1^n) - \mathrm{P}\left(\max_{1\le j\le p} Y_j \le t\right)\right|,$$
$$\rho_n^B := \sup_{t\in\mathbb{R}} \left| \mathrm{P}(\bar{W}\le t \mid X_1^n) - \mathrm{P}\left(\max_{1\le j\le p} Y_j \le t\right)\right|.$$

Observe that under the present assumptions, we may apply Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2014b) so that we have

$$\rho_n \le Cn^{-c}; \tag{62}$$

while applying Corollaries 4.1 and 4.2 in Chernozhukov, Chetverikov, and Kato (2014b) to the MB and EB procedures, respectively, we have for some $\nu_n := Cn^{-c}$,

$$\mathrm{P}(\rho_n^B < \nu_n) \ge 1 - Cn^{-c}. \tag{63}$$

We divide the rest of the proof into three steps. Step 1 establishes a relation between $c^B(\cdot)$ and $c_0(\cdot)$. Step 2 proves the assertion of the theorem. Step 3 provides auxiliary calculations. In particular, Step 3 shows that for some ζ_{n1} and ζ_{n2} satisfying $\zeta_{n1}\sqrt{\log p}+\zeta_{n2}\le Cn^{-c}$, we have

$$\mathrm{P}(|\bar{T}-T_0|>\zeta_{n1}) \le Cn^{-c}, \tag{64}$$
$$\mathrm{P}(\mathrm{P}(|W-\bar{W}|>\zeta_{n1}\mid X_1^n)>\zeta_{n2}) \le Cn^{-c}. \tag{65}$$

Step 1. We wish to prove that

$$\mathrm{P}(c^B(\alpha)\ge c_0(\alpha+\zeta_{n2}+\nu_n+8\zeta_{n1}\sqrt{\log p}))\ge 1-Cn^{-c}, \tag{66}$$
$$\mathrm{P}(c^B(\alpha)\le c_0(\alpha-\zeta_{n2}-\nu_n-8\zeta_{n1}\sqrt{\log p}))\ge 1-Cn^{-c}. \tag{67}$$

To establish (66), observe that for any $t\in\mathbb{R}$,

$$\mathrm{P}(W\le t\mid X_1^n) \le \mathrm{P}(\bar{W}\le t+\zeta_{n1}\mid X_1^n) + \mathrm{P}(|W-\bar{W}|>\zeta_{n1}\mid X_1^n) \tag{68}$$
$$\le \mathrm{P}\left(\max_{1\le j\le p} Y_j \le t+\zeta_{n1}\right) + \rho_n^B + \mathrm{P}(|W-\bar{W}|>\zeta_{n1}\mid X_1^n). \tag{69}$$

By Lemma A.4, for any $\gamma \in (0, 1-8\zeta_{n1}\sqrt{\log p})$ (note that $1-8\zeta_{n1}\sqrt{\log p} > 0$ for sufficiently large n),

$$\begin{aligned}
&\mathrm{P}\left(\max_{1\le j\le p} Y_j \le c_0(\gamma + 8\zeta_{n1}\sqrt{\log p}) + \zeta_{n1}\right) \\
&\quad \le \mathrm{P}\left(\max_{1\le j\le p} Y_j \le c_0(\gamma + 8\zeta_{n1}\sqrt{\log p})\right) + 2\zeta_{n1}(\sqrt{2\log p} + 1) \\
&\quad \le \mathrm{P}\left(\max_{1\le j\le p} Y_j \le c_0(\gamma + 8\zeta_{n1}\sqrt{\log p})\right) + 8\zeta_{n1}\sqrt{\log p} \\
&\quad = 1 - \gamma - 8\zeta_{n1}\sqrt{\log p} + 8\zeta_{n1}\sqrt{\log p} = 1 - \gamma,
\end{aligned}$$

where the third line follows from $p \ge 2$, so that $\sqrt{2\log p} \ge 1$, and the fourth line from the fact that the distribution of $\max_{1\le j\le p} Y_j$ has no point masses. Hence

$$c_0(\gamma + 8\zeta_{n1}\sqrt{\log p}) + \zeta_{n1} \le c_0(\gamma). \tag{70}$$

Therefore, setting $t = c_0(\alpha + \zeta_{n2} + \nu_n + 8\zeta_{n1}\sqrt{\log p})$ in (68)-(69), we obtain

$$\begin{aligned}
&\mathrm{P}(W \le c_0(\alpha + \zeta_{n2} + \nu_n + 8\zeta_{n1}\sqrt{\log p}) \mid X_1^n) \\
&\quad \le 1 - \alpha - \zeta_{n2} - \nu_n + \rho_n^B + \mathrm{P}(|W - \bar{W}| > \zeta_{n1} \mid X_1^n) < 1 - \alpha
\end{aligned}$$

on the event that $\rho_n^B < \nu_n$ and $\mathrm{P}(|W - \bar{W}| > \zeta_{n1} \mid X_1^n) \le \zeta_{n2}$, which holds with probability larger than $1 - Cn^{-c}$ by (63) and (65). This implies (66). By a similar argument, we can establish that (67) holds as well. This completes Step 1.

Step 2. Here we prove the asserted claims. Observe that under H_0,

$$\begin{aligned}
&\mathrm{P}(T > c^B(\alpha)) \le \mathrm{P}(\bar{T} > c^B(\alpha)) \\
&\quad \le \mathrm{P}(T_0 > c^B(\alpha) - \zeta_{n1}) + \mathrm{P}(|\bar{T} - T_0| > \zeta_{n1}) \\
&\quad \le \mathrm{P}(T_0 > c_0(\alpha + \zeta_{n2} + \nu_n + 8\zeta_{n1}\sqrt{\log p}) - \zeta_{n1}) + Cn^{-c} \\
&\quad \le \mathrm{P}(T_0 > c_0(\alpha + \zeta_{n2} + \nu_n + 16\zeta_{n1}\sqrt{\log p})) + Cn^{-c} \\
&\quad \le \mathrm{P}(\max_{1\le j\le p} Y_j > c_0(\alpha + \zeta_{n2} + \nu_n + 16\zeta_{n1}\sqrt{\log p})) + \rho_n + Cn^{-c} \\
&\quad = \alpha + \zeta_{n2} + \nu_n + 16\zeta_{n1}\sqrt{\log p} + \rho_n + Cn^{-c} \le \alpha + Cn^{-c},
\end{aligned}$$

where the third line follows from (64) and (66), the fourth line from (70), and the last line from (62) and construction of ν_n, ζ_{n1}, and ζ_{n2}. Hence, (30) follows. To prove (31), observe that when $\mu_j = 0$ for all $1 \le j \le p$, $T = \bar{T}$, and so

$$\begin{aligned}
\mathrm{P}(T > c^B(\alpha)) &= \mathrm{P}(\bar{T} > c^B(\alpha)) \\
&\geq \mathrm{P}(T_0 > c^B(\alpha) + \zeta_{n1}) - \mathrm{P}(|\bar{T} - T_0| > \zeta_{n1}) \\
&\geq \mathrm{P}(T_0 > c_0(\alpha - \zeta_{n2} - \nu_n - 8\zeta_{n1}\sqrt{\log p}) + \zeta_{n1}) - Cn^{-c} \\
&\geq \mathrm{P}(T_0 > c_0(\alpha - \zeta_{n2} - \nu_n - 16\zeta_{n1}\sqrt{\log p})) - Cn^{-c} \\
&\geq \mathrm{P}(\max_{1\leq j\leq p} Y_j > c_0(\alpha - \zeta_{n2} - \nu_n - 16\zeta_{n1}\sqrt{\log p})) - \rho_n - Cn^{-c} \\
&= \alpha - \zeta_{n2} - \nu_n - 16\zeta_{n1}\sqrt{\log p} - \rho_n - Cn^{-c} \geq \alpha - Cn^{-c},
\end{aligned}$$

where the third line follows from (64) and (67), the fourth line from (70), and the equality in the last line from the fact that the distribution of $\max_{1\leq j\leq p} Y_j$ has no point masses. Hence (31) follows. This completes Step 2.

Step 3. We wish to prove (64) and (65). 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 $B_n^2 \log^{7/2}(pn) \leq C_1 n^{1/2-c_1}$, these choices satisfy $\zeta_{n1}\sqrt{\log p} + \zeta_{n2} \leq Cn^{-c}$.

We first verify (64). 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

$$\mathrm{P}\left(\max_{1\leq j\leq p} |(\sigma_j/\widehat{\sigma}_j) - 1| > n^{-1/2+c_1/4} B_n^2 \log p\right) \leq Cn^{-c}. \tag{71}$$

Moreover, by Markov's inequality and (52),

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

Hence (64) 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 (65), let A_n be the event such that

$$A_n := \left\{\max_{1\leq j\leq p} |(\widehat{\sigma}_j/\sigma_j) - 1| \leq (n^{-1/2+c_1/4} B_n^2 \log p) \wedge (1/4)\right\}.$$

We have seen that $\mathrm{P}(A_n) > 1 - Cn^{-c}$. We consider MB and EB procedures separately.

Consider the MB procedure first, so that $W = W^{MB}$ and $\bar{W} = \bar{W}^{MB}$. Observe that

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

Conditional on the data X_1^n, the vector $(\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \widehat{\mu}_j)/\widehat{\sigma}_j])_{1\le j\le p}$ is normal with mean zero and all the diagonal elements of the covariance matrix are one. Hence $\mathrm{E}[|W^{MB}| \mid X_1^n] \le \sqrt{2\log(2p)}$, so that by Markov's inequality, on the event A_n,

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

which is bounded by $Cn^{-c_1/4}$, so that (65) for the MB procedure is verified.

Now consider the EB procedure. On the event $A_n \cap \{\mathrm{P}(|W^{MB} - \bar{W}^{MB}| > \zeta_{n1} \mid X_1^n) \le \zeta_{n2}\} \cap \{\rho_n^{MB} < \nu_n\} \cap \{\rho_n^{EB} < \nu_n\}$, which holds with probability larger than $1 - Cn^{-c}$,

$$\begin{aligned}
&\mathrm{P}(|W^{EB} - \bar{W}^{EB}| > \zeta_{n1} \mid X_1^n) \\
&\quad \le \mathrm{P}(\max_{1\le j\le p} |(\sigma_j/\widehat{\sigma}_j) - 1| \times |\bar{W}^{EB}| > \zeta_{n1} \mid X_1^n) \\
&\quad \le \mathrm{P}(\max_{1\le j\le p} |(\sigma_j/\widehat{\sigma}_j) - 1| \times |\bar{W}^{MB}| > \zeta_{n1} \mid X_1^n) + \rho_n^{EB} + \rho_n^{MB} \\
&\quad \le \mathrm{P}(\max_{1\le j\le p} |(\widehat{\sigma}_j/\sigma_j) - 1| \times |W^{MB}| > \zeta_{n1}/4 \mid X_1^n) + \rho_n^{EB} + \rho_n^{MB} \le Cn^{-c},
\end{aligned}$$

so that (65) for the EB procedure is verified. This completes the proof. □

A.5. **Proof of Theorem 4.4.** Here c, C denote generic positive constants depending only on c_1, C_1; their values may change from place to place. Let $\widehat{J}_B$ stay either for $\widehat{J}_{MB}$ or $\widehat{J}_{EB}$ depending on which bootstrap procedure is used. 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 $\max_{1\le j\le p} Y_j$. Recall that in the proof of Theorem 4.3, we established that with probability larger than $1 - Cn^{-c}$, $c^B(\alpha) \ge c_0(\alpha + \varphi_n)$ and $c^B(\alpha) \le c_0(\alpha - \varphi_n)$ for some $0 < \varphi_n \le Cn^{-c}$; see (66) and (67). Define

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

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 $j \in J_2^c$.

Like in the proof of Theorem 4.2, observe that

$$\widehat{\mu}_j > 0 \text{ for some } j \in J_2^c \Rightarrow \max_{1\le j\le p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\sigma_j > c_0(\beta_n + \varphi_n),$$

so that it is enough to prove that

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

But this follows from Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2014b) (and the fact that $\varphi_n = C' n^{-c'}$). This concludes Step 1.

Step 2. We wish to prove that with probability larger than $1 - \beta_n - Cn^{-c}$, $\widehat{J}_B \supset J_2$.

Like in the proof of Theorem 4.2, observe that

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

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

$$\begin{aligned}&2\widehat{\sigma}_j c^B(\beta_n)-\sigma_j c_0(\beta_n+\varphi_n)\ge(2\widehat{\sigma}_j-\sigma_j)c_0(\beta_n+\varphi_n)\\&\quad=\sigma_j(1+2(\widehat{\sigma}_j/\sigma_j-1))c_0(\beta_n+\varphi_n)\ge(1-r)\sigma_j c_0(\beta_n+\varphi_n),\end{aligned}$$

we have

$$\begin{aligned}\mathrm{P}(\widehat{J}_B \not\supset J_2)\le \mathrm{P}\left(\max_{1\le j\le p}\frac{\sqrt{n}(\mu_j-\widehat{\mu}_j)}{\sigma_j}>(1-r)c_0(\beta_n+\varphi_n)\right)&\qquad(72)\\+\mathrm{P}\left(c^B(\beta_n)<c_0(\beta_n+\varphi_n)\right)+\mathrm{P}\left(\max_{1\le j\le p}|(\widehat{\sigma}_j/\sigma_j)-1|>r/2\right).&\end{aligned}$$

By Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2014b), the probability on the right-hand side of (72) is bounded by

$$\mathrm{P}\left(\max_{1\le j\le p}Y_j>(1-r)c_0(\beta_n+\varphi_n)\right)+Cn^{-c}.$$

Moreover, by Lemma A.4,

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

which is bounded by $\beta_n+\varphi_n+Cr\log(pn)$. Thus,

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

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-hand side of the inequality above is bounded by Cn^{-c}, and

$$r\log(pn)\le n^{-(1-c_1)/2}B_n^2\log^2(pn)\le C_1n^{-c_1/2},$$

because of the assumption that $B_n^2\log^{7/2}(pn)\le C_1n^{1/2-c_1}$. This leads to the conclusion of Step 2.

Step 3. We are now in position to finish the proof of the theorem. Assume first that $J_2=\emptyset$. Then by Step 1 we have that $T\le 0$ with probability larger than $1-\beta_n-Cn^{-c}$. But as $c^{B,2S}(\alpha)\ge 0$ (recall that $\alpha<1/2$), we have $\mathrm{P}(T>c^{B,2S}(\alpha))\le\beta_n+Cn^{-c}\le\alpha+Cn^{-c}$. Now consider the case where $J_2\ne\emptyset$. Define $c^{B,2S}(\alpha,J_2)$ by the same bootstrap procedure as $c^{B,2S}(\alpha)$

with $\widehat{J}_B$ replaced by J_2. Note that $c^{B,2S}(\alpha) \geq c^{B,2S}(\alpha, J_2)$ on the event $\widehat{J}_B \supset J_2$. Therefore, arguing as in Step 3 of the proof of Theorem 4.2,

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

This gives the first assertion of the theorem.

Moreover, when $\mu_j = 0$ for all $1 \leq j \leq p$, we have $J_2 = \{1, \dots, p\}$. Hence by Step 2, $c^{B,2S}(\alpha) = c^{B,2S}(\alpha, J_2)$ with probability larger than $1 - \beta_n - Cn^{-c}$. Therefore,

$$\begin{aligned}
\mathrm{P}(T > c^{B,2S}(\alpha)) &= \mathrm{P}\left(\max_{1 \leq j \leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j > c^{B,2S}(\alpha)\right) \\
&\geq \mathrm{P}\left(\max_{1 \leq j \leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j > c^{B,2S}(\alpha, J_2)\right) - \beta_n - Cn^{-c} \\
&\geq \alpha - 3\beta_n - Cn^{-c}.
\end{aligned}$$

This gives the second assertion of the theorem. Finally, the last assertion follows trivially. This completes the proof of the theorem. □

A.6. **Proof of Theorem 4.5.** Recall the set $J_1 \subset \{1, \dots, p\}$ defined in (54). By Steps 1 and 2 in the proof of Theorem 4.2, we see that

$$\begin{aligned}
&\mathrm{P}(\widehat{\mu}_j \leq 0 \text{ for all } j \in J_1^c) > 1 - \beta_n - Cn^{-c}, \\
&\mathrm{P}(\widehat{J}_{SN} \supset J_1) > 1 - \beta_n - Cn^{-c},
\end{aligned}$$

where c, C are some positive constants depending only on c_1, C_1. The rest of the proof is completely analogous to Step 3 in the proof of Theorem 4.4 and hence omitted. □

A.7. **Proof of Theorem 4.6.** Here c, C denote generic positive constants depending only on c_1, C_1, c_2, C_2; their values may change from place to place. Define

$$\begin{aligned}
&J_2 := \left\{j \in \{1, \dots, p\} : \sqrt{n}\mu_j/\sigma_j > -c_0(\beta_n + \varphi_n)\right\}, \ J_2^c := \{1, \dots, p\} \backslash J_2, \\
&J_3 := \left\{j \in \{1, \dots, p\} : \sqrt{n}\mu_{jl}^V/\sigma_{jl}^V > -2c_0^V(\beta_n) \text{ for all } l = 1, \dots, r\right\}
\end{aligned}$$

where $c^0(\beta_n + \varphi_n)$ is defined as in the proof of Theorem 4.4 and $c_0^V(\beta_n)$ is the $(1 - \beta_n)$-quantile of the distribution of $\max_{j,l} Y_{jl}^V$ where $\{Y_{jl}^V, 1 \leq j \leq p, 1 \leq l \leq r\}$ is a sequence of Gaussian random variables with mean zero and covariance $\mathrm{E}[Y_{jl}^V Y_{j'l'}^V] = \mathrm{E}[Z_{1jl}^V Z_{1j'l'}^V]$.

By the same arguments as those used in Steps 1 and 2 of the proof of Theorem 4.4, we have

$$\begin{aligned}
&\mathrm{P}(J_2 \subset \widehat{J}_B) \geq 1 - \beta_n - Cn^{-c},\\
&\mathrm{P}(J_3 \subset \widehat{J}''_B) \geq 1 - \beta_n - Cn^{-c},\\
&\mathrm{P}(\widehat{J}'_B \subset J_3) \geq 1 - \beta_n - Cn^{-c},\\
&\mathrm{P}(\widehat{\mu}_j \leq 0, \text{ for all } j \in J_2^c) \geq 1 - \beta_n - Cn^{-c}.
\end{aligned}$$

Define $c^{B,3S}(\alpha, J_2 \cap J_3)$ by the same bootstrap procedure as $c^{B,3S}(\alpha)$ with $\widehat{J}_B \cap \widehat{J}''_B$ replaced by $J_2 \cap J_3$. Then inequalities above imply that $c^{B,3S}(\alpha, J_2 \cap J_3) \leq c^{B,3S}(\alpha)$ with probability larger than $1 - 2\beta_n - Cn^{-c}$. Therefore, by an argument similar to that used in Step 3 of the proof of Theorem 4.4, with maximum over empty set understood as 0, we have

$$\begin{aligned}
\mathrm{P}(T > c^{B,3S}(\alpha)) &\leq \mathrm{P}\left(\max_{j \in J_2 \cap \widehat{J}'_B} \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j > c^{B,3S}(\alpha)\right) + \beta_n + Cn^{-c}\\
&\leq \mathrm{P}\left(\max_{j \in J_2 \cap \widehat{J}'_B} \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j > c^{B,3S}(\alpha, J_2 \cap J_3)\right) + 3\beta_n + Cn^{-c}\\
&\leq \mathrm{P}\left(\max_{j \in J_2 \cap J_3} \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j > c^{B,3S}(\alpha, J_2 \cap J_3)\right) + 4\beta_n + Cn^{-c}\\
&\leq \alpha - 4\beta_n + 4\beta_n + Cn^{-c} = \alpha + Cn^{-c}.
\end{aligned}$$

This completes the proof of the theorem. □

A.8. **Proof of Lemma 5.1.** Denote $X_1^n = (X_1, \dots, X_n)$. Let $\phi_n : \mathbb{R}^{pn} \to [0,1], X_1^n \mapsto \phi_n(X_1^n)$, be a test such that $\mathrm{E}_\mu[\phi_n(X_1^n)] \leq \alpha$ 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 equals $r\sigma_j$. 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

$$\inf_{\max_{1 \leq j \leq p}(\mu_j/\sigma_j) \geq r} \mathrm{E}_\mu[\phi_n(X_1^n)] - \alpha \leq \frac{1}{p}\sum_{j=1}^p \mathrm{E}_j[\phi_n(X_1^n)] - \mathrm{E}_0[\phi_n(X_1^n)]. \quad (73)$$

Further,

$$\mathrm{E}_j[\phi(X_1^n)] = \mathrm{E}_0[e^{nr\widehat{\mu}_j/\sigma_j - nr^2/2}\phi_n(X_1^n)]$$

where $\widehat{\mu}_j = \mathbb{E}_n[X_{ij}]$. Hence the right-hand side of (73) is written as

$$\mathrm{E}_0\left[\left\{\frac{1}{p}\sum_{j=1}^p e^{nr\widehat{\mu}_j/\sigma_j - nr^2/2} - 1\right\}\phi_n(X_1^n)\right] \leq \mathrm{E}_0\left[\left|\frac{1}{p}\sum_{j=1}^p e^{nr\widehat{\mu}_j/\sigma_j - nr^2/2} - 1\right|\right].$$

Note that under $\mu = 0$, $\sqrt{n}\widehat{\mu}_1/\sigma_1, \dots, \sqrt{n}\widehat{\mu}_p/\sigma_p \sim N(0,1)$ i.i.d. Hence we obtain the assertion (43).

The second assertion follows from application of Lemma 6.2 in Dümbgen and Spokoiny (2001). This completes the proof of the lemma. □

A.9. **Proof of Lemma 5.2.** Let $j^* \in \{1, \ldots, p\}$ be any index such that $\mu_{j^*}/\sigma_{j^*} = \max_{1 \leq j \leq p}(\mu_j/\sigma_j)$. Let $A_{n,1}$ and $A_{n,2}$ be the events that $|\widehat{\sigma}_{j^*}/\sigma_{j^*} - 1| \leq \delta$ and $\widehat{c}(\alpha) \leq (1+\underline{\epsilon})\sqrt{2\log(p/\alpha)}$, respectively. Then on the event $A_{n,1} \cap A_{n,2}$,

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

so that

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

Hence we have

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

By Markov's inequality, we have

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

This completes the proof. □

A.10. **Proof of Corollary 5.1.** 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 $M_{n,4}^2 \log^{1/2} p \leq C_1 n^{1/2-c_1}$, by Markov's inequality, there exists $\delta_n \leq \min\{C\log^{-1/2} p, 1/2\}$ such that

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

Hence, by Lemma 5.2, we only have to verify that

$$\mathrm{P}(\widehat{c}(\alpha) > (1 + C\log^{-1/2} p)\sqrt{2\log(p/\alpha)}) \leq Cn^{-c}. \tag{74}$$

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

$$c^{SN,2S}(\alpha) \leq c^{SN}(\alpha/3), \ c^{B,2S}(\alpha) \vee c^{B,H}(\alpha) \leq c^{B}(\alpha/3)$$

where $B = MB$ or EB, so that it suffices to verify (74) with $\widehat{c}(\alpha) = c^{SN}(\alpha)$, $c^{MB}(\alpha)$, and $c^{EB}(\alpha)$.

For $\widehat{c}(\alpha) = c^{SN}(\alpha)$, since $\Phi^{-1}(1-p/\alpha) \leq \sqrt{2\log(p/\alpha)}$ and $\log^{3/2} p \leq C_1 n$, it is straightforward to see that (74) 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 (74) can be verified by simple algebra.

Now consider $\widehat{c}(\alpha) = c^{EB}(\alpha)$. It is established in Step 1 of the proof of Theorem 4.3 that there exists a sequence $\varphi_n \geq 0$ such that $\varphi_n \leq Cn^{-c}$ and $\mathrm{P}(c^{EB}(\alpha) > c_0(\alpha - \varphi_n)) \leq Cn^{-c}$ where $c_0(\alpha - \varphi_n)$ is the $(1 - \alpha + \varphi_n)$th quantile of the distribution of $\max_{1\leq j\leq p} Y_j$ and $(Y_1, \dots, Y_p)^T$ is a normal vector with mean zero and all diagonal elements of the covariance matrix equal to one. By Lemma A.4,

$$c_0(\alpha - \varphi_n) \leq \sqrt{2\log p} + \sqrt{2\log(1/(\alpha - \varphi_n))}.$$

In addition, simple algebra shows that

$$(1 + C\log^{-1/2} p)\sqrt{2\log(p/\alpha)} > \sqrt{2\log p} + \sqrt{2\log(1/(\alpha - \varphi_n))}$$

if C is chosen sufficiently large (and depending on α). Combining these inequalities gives (74). This completes the proof. □

A.11. **Proof of Theorem 6.1.** The theorem readily follows from Theorems 4.1-4.5. □

A.12. **Proof of Theorem 7.1.** Here c, c', C, C' denote generic positive constants depending only on c_1, c_2, C_1; their values may change from place to place. It suffices to show that $|\mathrm{P}(\check{T} \leq \widehat{c}^{BMB}(\alpha)) - \alpha| \leq Cn^{-c}$ when $\mu_j = 0, 1 \leq \forall j \leq p$. Suppose that $\mu_j = 0, 1 \leq \forall j \leq p$. We use the extensions of the results in Chernozhukov, Chetverikov, and Kato (2013a) to dependent data proved in Appendix B ahead. Note that since $\log(pn) \leq C\sqrt{q}$ (which follows from $(r/q)\log^2 p \leq C_1 n^{-c_2}$), $\sqrt{q} D_n \log^{7/2}(pn) \leq CqD_n \log^{5/2}(pn) \leq C' n^{1/2-c_2}$, so that by Theorem B.1 in Appendix B,

$$\sup_{t\in\mathbb{R}} |\mathrm{P}(\check{T} \leq t) - \mathrm{P}(\max_{1\leq j\leq p} \check{Y}_j \leq t)| \leq Cn^{-c}, \tag{75}$$

where $\check{Y} = (\check{Y}_1, \dots, \check{Y}_p)^T$ is a centered normal random vector with covariance matrix $\mathrm{E}[\check{Y}\check{Y}^T] = (1/(mq))\sum_{l=1}^m \mathrm{E}[(\sum_{i\in I_l} X_i)(\sum_{i\in I_l} X_i)^T]$. Note that $c_1 \leq \underline{\sigma}^2(q) \leq \mathrm{E}[\check{Y}_j^2] \leq \overline{\sigma}^2(q) \leq C_1, 1 \leq \forall j \leq p$.

Let $\check{W}_0 = \max_{1\leq j\leq p}(1/\sqrt{mq})\sum_{l=1}^m \epsilon_l \sum_{i\in I_l} X_{ij}$. Then by Theorem B.2, with probability larger than $1 - Cn^{-c}$,

$$\sup_{t\in\mathbb{R}} |\mathrm{P}(\check{W}_0 \leq t \mid X_1^n) - \mathrm{P}(\max_{1\leq j\leq p} \check{Y}_j \leq t)| \leq C'n^{-c'}.$$

Observe that $|\check{W} - \check{W}_0| \leq \max_{1\leq j\leq p} |\sqrt{n}\widehat{\mu}_j| \cdot |m^{-1}\sum_{l=1}^m \epsilon_l|$. Here since $q \leq Cn^{1/2-c_2}$, we have $m \geq n/(4q) \geq C^{-1}n^{1/2-c_2}$, so that by Markov's inequality, $\mathrm{P}(|m^{-1}\sum_{l=1}^m \epsilon_l| > Cn^{-1/4+5c_2/8}) \leq n^{-c_2/8}$. On the other hand, by applying Theorem B.1 to $(X_{i1}, \dots, X_{ip}, -X_{i1}, \dots, -X_{ip})^T$, we have

$$\sup_{t\in\mathbb{R}} |\mathrm{P}(\max_{1\leq j\leq p} |\sqrt{n}\widehat{\mu}_j| \leq t) - \mathrm{P}(\max_{1\leq j\leq p} |\check{Y}_j| \leq t)| \leq Cn^{-c}.$$

Since $\mathrm{E}[\max_{1\le j\le p}|\check{Y}_j|]\le C\sqrt{\log p}$, we conclude that

$$\mathrm{P}(\max_{1\le j\le p}|\sqrt{n}\widehat{\mu}_j|>Cn^{c_2/8}\sqrt{\log p})\le C'n^{-c}.$$

Hence with probability larger than $1-Cn^{-c}$,

$$\mathrm{P}(|\check{W}-\check{W}_0|>\zeta_n\mid X_1^n)\le n^{-c'},$$

where $\zeta_n=C'n^{-1/4+3c_2/4}\sqrt{\log p}$. Note that since $qD_n\log^{5/2}(pn)\le C_1n^{1/2-c_2}$, $n^{-1/4+c_2/2}\log p\le Cq^{-1/2}\le C'n^{-c_2/2}$ (the second inequality follows from $(r/q)\log^2 p\le C_1n^{-c_2}$ so that $q^{-1}\le Cn^{-c_2}$), and hence $\zeta_n\sqrt{\log p}\le Cn^{-c_2/4}$. Using the anti-concentration property of $\max_{1\le j\le p}\check{Y}_j$ (see Step 3 in the proof of Theorem B.1), we conclude that with probability larger than $1-Cn^{-c}$,

$$\sup_{t\in\mathbb{R}}|\mathrm{P}(\check{W}\le t\mid X_1^n)-\mathrm{P}(\max_{1\le j\le p}\check{Y}_j\le t)|\le C'n^{-c'}.$$

The desired assertion follows from combining this inequality with (75). □

A.13. **Proof of Theorem 7.2.** Here c,C denote generic positive constants depending only on c_1,C_1; their values may change from place to place. Define

$$\bar{T}:=\max_{1\le j\le p}\frac{\sqrt{n}(\widehat{\mu}_j-\mu_j)}{\widehat{\sigma}_j},\ T_0:=\max_{1\le j\le p}\frac{\sqrt{n}(\widehat{\mu}_{j,0}-\mu_j)}{\sigma_j},$$
$$W^{MB}:=\max_{1\le j\le p}\frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(\widehat{X}_{ij}-\widehat{\mu}_j)]}{\widehat{\sigma}_j},\ \bar{W}^{MB}:=\max_{1\le j\le p}\frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_{j,0})]}{\sigma_j},$$
$$W^{EB}:=\max_{1\le j\le p}\frac{\sqrt{n}\mathbb{E}_n[\widehat{X}^*_{ij}-\widehat{\mu}_j]}{\widehat{\sigma}_j},\ \bar{W}^{EB}:=\max_{1\le j\le p}\frac{\sqrt{n}\mathbb{E}_n[X^*_{ij}-\widehat{\mu}_{j,0}]}{\sigma_j},$$

where $\widehat{X}^*_1,\dots,\widehat{X}^*_n$ is an empirical bootstrap sample from $\widehat{X}_1,\dots,\widehat{X}_n$, and $X^*_1,\dots,X^*_n$ is an empirical bootstrap sample from $X_1,\dots,X_n$. Observe that the critical values $c^{MB,2S}(\alpha)$ and $c^{EB,2S}(\alpha)$ are based on the bootstrap statistics W^{MB} and W^{EB}.

We divide the proof into several steps. In Steps 1, 2, and 3, we prove that

$$\mathrm{P}\left(|\bar{T}-T_0|>\zeta'_{n1}\right)\le Cn^{-c}, \tag{76}$$
$$\mathrm{P}(\mathrm{P}(|W^{MB}-\bar{W}^{MB}|>\zeta'_{n1}\mid X_1^n)>Cn^{-c})\le Cn^{-c}, \tag{77}$$
$$\mathrm{P}(\mathrm{P}(|W^{EB}-\bar{W}^{EB}|>\zeta'_{n1}\mid X_1^n)>Cn^{-c})\le Cn^{-c}, \tag{78}$$

respectively, for some ζ'_{n1} satisfying $\zeta'_{n1}\sqrt{\log p}\le Cn^{-c}$. In Step 4, we prove an auxiliary result that

$$\mathrm{P}\left(\max_{1\le j\le p}|1-\widehat{\sigma}_j/\widehat{\sigma}_{j,0}|>C\zeta_{n1}\right)\le Cn^{-c}. \tag{79}$$

Given results (76)-(78), the conclusions of the theorem follow by repeating the arguments used in the proofs of Theorems 4.3 and 4.4.

In the proof, we will frequently use the following implications of Lemma A.5 (recall that $\widehat{\sigma}_j$ in Lemma A.5 is denoted as $\widehat{\sigma}_{j,0}$ in this proof):

$$\mathrm{P}\left(\max_{1\leq j\leq p}(\sigma_j/\widehat{\sigma}_{j,0})^2>2\right)\leq Cn^{-c}, \tag{80}$$

$$\mathrm{P}\left(\max_{1\leq j\leq p}(\widehat{\sigma}_{j,0}/\sigma_j)^2>2\right)\leq Cn^{-c}. \tag{81}$$

Step 1. Here we wish to prove (76). Define $T_0' := \max_{1\leq j\leq p}\sqrt{n}(\widehat{\mu}_{j,0}-\mu_j)/\widehat{\sigma}_j$. Observe that

$$\begin{aligned}|\bar{T}-T_0'| &\leq \max_{1\leq j\leq p}\left|\frac{\sqrt{n}(\widehat{\mu}_j-\widehat{\mu}_{j,0})}{\widehat{\sigma}_j}\right| \leq C\max_{1\leq j\leq p}\left|\frac{\sqrt{n}(\widehat{\mu}_j-\widehat{\mu}_{j,0})}{\sigma_j}\right| \\ &\leq C\max_{1\leq j\leq p}|\sqrt{n}(\widehat{\mu}_j-\widehat{\mu}_{j,0})|\leq C\zeta_{n1}\end{aligned}$$

with probability larger than $1-Cn^{-c}$ where the second inequality in the first line follows from (79) and (80) and the second line follows from assumptions. Also,

$$|T_0'-T_0|\leq\max_{1\leq j\leq p}|\sigma_j/\widehat{\sigma}_j-1|\times\max_{1\leq j\leq p}|\sqrt{n}\mathbb{E}_n[Z_{ij}]|,$$

where $Z_{ij}=(X_{ij}-\mu_j)/\sigma_j$. As shown in Step 3 of the proof of Theorem 4.3,

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

In addition, using an elementary inequality $|ab-1|\leq|a||b-1|+|a-1|$ with $a=\sigma_j/\widehat{\sigma}_{j,0}$ and $b=\widehat{\sigma}_{j,0}/\widehat{\sigma}_j$, we obtain from (71) in the proof of Theorem 4.3, (79), and (80) that

$$\mathrm{P}\left(\max_{1\leq j\leq p}|\sigma_j/\widehat{\sigma}_j-1|>C(n^{-1/2+c_1/4}B_n^2\log p+\zeta_{n1})\right)\leq Cn^{-c}$$

(remember that $\widehat{\sigma}_j$ in the proof of Theorem 4.3 corresponds to $\widehat{\sigma}_{j,0}$ here). Therefore, the claim of this step holds with $\zeta_{n1}' := C(n^{-1/2+c_1/2}B_n^2(\log p)^{3/2}+\zeta_{n1}n^{c_1/4}\sqrt{\log p})$ for sufficiently large C.

Step 2. Here we wish to prove (77). Let $\widehat{W}^{MB} := \max_{1\leq j\leq p}\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_{j,0})]/\widehat{\sigma}_j$. By (79) and (80), with probability larger than $1-Cn^{-c}$,

$$\begin{aligned}|W^{MB}-\widehat{W}^{MB}| &\leq \max_{1\leq j\leq p}\frac{|\sqrt{n}\mathbb{E}_n[\epsilon_i(\widehat{X}_{ij}-X_{ij}-\widehat{\mu}_j+\widehat{\mu}_{j,0})]|}{\widehat{\sigma}_j} \\ &\leq C\max_{1\leq j\leq p}\frac{|\sqrt{n}\mathbb{E}_n[\epsilon_i(\widehat{X}_{ij}-X_{ij}-\widehat{\mu}_j+\widehat{\mu}_{j,0})]|}{\sigma_j} \\ &\leq C\max_{1\leq j\leq p}|\sqrt{n}\mathbb{E}_n[\epsilon_i(\widehat{X}_{ij}-X_{ij}-\widehat{\mu}_j+\widehat{\mu}_{j,0})]|,\end{aligned}$$

where the third inequality follows from the assumption that $\sigma_j\geq c_1$ for all $j=1,\dots,p$. Conditional on X_1^n, the vector $(\sqrt{n}\mathbb{E}_n[\epsilon_i(\widehat{X}_{ij}-X_{ij}-\widehat{\mu}_j+$

$\widehat{\mu}_{j,0})])_{1\le j\le p}$ is normal with mean zero and all diagonal elements of the covariance matrix bounded by $\max_{1\le j\le p}\mathbb{E}_n[(\widehat{X}_{ij}-X_{ij}-\widehat{\mu}_j+\widehat{\mu}_{j,0})^2]$. As established in in Step 4 below, the last quantity is bounded by $C\zeta_{n1}^2$ with probability larger than $1-Cn^{-c}$. Therefore,

$$\mathrm{P}(\mathrm{P}(|W^{MB}-\widehat{W}^{MB}|>C\zeta_{n1}\sqrt{\log p}\mid X_1^n)>Cn^{-c})\le Cn^{-c}. \tag{82}$$

Moreover

$$|\widehat{W}^{MB}-\bar{W}^{MB}|\le\max_{1\le j\le p}|\sigma_j/\widehat{\sigma}_j-1|\times\bar{W}^{MB}.$$

Now observe that $\bar{W}^{MB}=\max_{1\le j\le p}\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_{j,0})/\sigma_j]$ and conditional on the data X_1^n, the vector $(\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\widehat{\mu}_{j,0})/\sigma_j])_{1\le j\le p}$ is normal with mean zero and all diagonal elements of the covariance matrix bounded by $\max_{1\le j\le p}(\widehat{\sigma}_{j,0}^2/\sigma_j^2)$. By (81), the last quantity is bounded by 2 with probability larger than $1-Cn^{-c}$. Therefore,

$$\mathrm{P}(\mathrm{P}(|\widehat{W}^{MB}-\bar{W}^{MB}|>\zeta_{n1}'\mid X_1^n)>Cn^{-c})\le Cn^{-c} \tag{83}$$

where ζ_{n1}' is defined in Step 1. Combining (82) and (83) leads to the assertion of this step.

Step 3. Here we wish to prove (78). Let $\widehat{W}^{EB}:=\max_{1\le j\le p}\sqrt{n}\mathbb{E}_n[X_{ij}^*-\widehat{\mu}_{j,0}]/\widehat{\sigma}_j$. By (79) and (80), with probability larger than $1-Cn^{-c}$,

$$\begin{aligned}|W^{EB}-\widehat{W}^{EB}|&\le\max_{1\le j\le p}\frac{|\sqrt{n}\mathbb{E}_n[\widehat{X}_{ij}^*-X_{ij}^*-\widehat{\mu}_j+\widehat{\mu}_{j,0}]|}{\widehat{\sigma}_j}\\&\le C\max_{1\le j\le p}\frac{|\sqrt{n}\mathbb{E}_n[\widehat{X}_{ij}^*-X_{ij}^*-\widehat{\mu}_j+\widehat{\mu}_{j,0}]|}{\sigma_j}\\&\le C\max_{1\le j\le p}|\sqrt{n}\mathbb{E}_n[\widehat{X}_{ij}^*-X_{ij}^*-\widehat{\mu}_j+\widehat{\mu}_{j,0}]|,\end{aligned}$$

where the third inequality follows from the assumption that $\sigma_j\ge c_1$ for all $1\le j\le p$. Applying Lemma A.3 conditional on the data X_1^n, we have

$$\begin{aligned}&\mathrm{E}\left[\max_{1\le j\le p}|\sqrt{n}\mathbb{E}_n[\widehat{X}_{ij}^*-X_{ij}^*-\widehat{\mu}_j+\widehat{\mu}_{j,0}]|\mid X_1^n\right]\\&\quad\le C\left(\max_{1\le j\le p}(\mathbb{E}_n[(\widehat{X}_{ij}-X_{ij})^2]\log p)^{1/2}+\max_{i,j}|\widehat{X}_{ij}-X_{ij}|(\log p)/\sqrt{n}\right).\end{aligned}$$

Therefore, by Markov's inequality, we have

$$\mathrm{P}(\mathrm{P}(|W^{EB}-\widehat{W}^{EB}|>C\zeta_{n1}n^{c_1/4}\sqrt{\log p}\mid X_1^n)>Cn^{-c})\le Cn^{-c}. \tag{84}$$

Moreover

$$|\widehat{W}^{EB}-\bar{W}^{EB}|\le\max_{1\le j\le p}|\sigma_j/\widehat{\sigma}_j-1|\times\bar{W}^{EB}.$$

Applying Lemma A.3 conditional on the data X_1^n once again, we have

$$\mathrm{E}[\bar{W}^{EB}\mid X_1^n]\le C\left(\max_{1\le j\le p}(\widehat{\sigma}_{j,0}/\sigma_j)+\max_{i,j}\frac{|X_{ij}-\mu_j|}{\sigma_j}(\log p)/\sqrt{n}\right).$$

By (81), $\max_{1\le j\le p}(\widehat{\sigma}_{j,0}/\sigma_j) \le \sqrt{2}$ with probability larger than $1-Cn^{-c}$. Here for $Z_{ij} = (X_{ij}-\mu_j)/\sigma_j$,

$$\mathrm{E}\left[\max_{1\le j\le p}|Z_{ij}|\right] \le \left(\mathrm{E}\left[\max_{i,j}|Z_{ij}|^4\right]\right)^{1/4} \le \left(\mathrm{E}\left[n\max_{1\le j\le p}|Z_{ij}|^4\right]\right)^{1/4} = n^{1/4}B_n.$$

Hence, by Markov's inequality and the assumption that $B_n^2\log^{7/2}(pn) \le C_1 n^{1/2-c_1}$, we have $\max_{i,j}(|X_{ij}-\mu_j|/\sigma_j)(\log p)/\sqrt{n} \le C\sqrt{\log p}$ with probability larger than $1-Cn^{-c}$ for sufficiently large C. Therefore,

$$\mathrm{P}(\mathrm{P}(|\widehat{W}^{EB}-\bar{W}^{EB}| > C\zeta_{n1}\sqrt{\log p} \mid X_1^n) > Cn^{-c}) \le Cn^{-c}. \tag{85}$$

Combining (84) and (85) leads to the assertion of this step.

Step 4. Here we wish to prove (79). Using (80), we obtain that with probability larger than $1-Cn^{-c}$, for all $j=1,\dots,p$,

$$\begin{aligned}\left|1-\frac{\widehat{\sigma}_j}{\widehat{\sigma}_{j,0}}\right| &\le \left|1-\left(\frac{\widehat{\sigma}_j}{\widehat{\sigma}_{j,0}}\right)^2\right| = \frac{1}{\widehat{\sigma}_{j,0}^2}\left|\widehat{\sigma}_j^2-\widehat{\sigma}_{j,0}^2\right| \le \frac{2}{\sigma_j^2}\left|\widehat{\sigma}_j^2-\widehat{\sigma}_{j,0}^2\right| \\ &= \frac{2}{\sigma_j^2}\left|\mathbb{E}_n[(\widehat{X}_{ij}-\widehat{\mu}_j)^2-(X_{ij}-\widehat{\mu}_{j,0})^2]\right|.\end{aligned}$$

Since $a^2-b^2=(a-b)^2+2b(a-b)$ for any $a,b\in\mathbb{R}$, we have, by the Cauchy-Schwarz inequality,

$$\begin{aligned}|\mathbb{E}_n[(\widehat{X}_{ij}-\widehat{\mu}_j)^2-(X_{ij}-\widehat{\mu}_{j,0})^2]| &\le \mathbb{E}_n[(\widehat{X}_{ij}-X_{ij}-\widehat{\mu}_j+\widehat{\mu}_{j,0})^2] \\ &\quad + 2\widehat{\sigma}_{j,0}\left(\mathbb{E}_n[(\widehat{X}_{ij}-X_{ij}-\widehat{\mu}_j+\widehat{\mu}_{j,0})^2]\right)^{1/2}.\end{aligned}$$

Also,

$$\left(\mathbb{E}_n[(\widehat{X}_{ij}-X_{ij}-\widehat{\mu}_j+\widehat{\mu}_{j,0})^2]\right)^{1/2} \le (\mathbb{E}_n[(\widehat{X}_{ij}-X_{ij})^2])^{1/2}+|\widehat{\mu}_j-\widehat{\mu}_{j,0}|,$$

which is further bounded by $C\zeta_{n1}$ with probability larger than $1-Cn^{-c}$. Taking these inequalities together, we conclude that with probability larger than $1-Cn^{-c}$, for all $j=1,\dots,p$,

$$\left|1-\frac{\widehat{\sigma}_j}{\widehat{\sigma}_{j,0}}\right| \le \frac{2(C\zeta_{n1})^2}{\sigma_j^2}+\frac{4\widehat{\sigma}_{j,0}C\zeta_{n1}}{\sigma_j^2} \le C\zeta_{n1},$$

where the last inequality follows from the assumption that $\sigma_j \ge c_1$ for all $j=1,\dots,p$ and inequality (81). This leads to the assertion of Step 4 and completes the proof of the theorem. □

Appendix B. High dimensional CLT under dependence

In this appendix, we extend the results of Chernozhukov, Chetverikov, and Kato (2013a) to dependent data. Let $X_1,\dots,X_n$ be possibly dependent random vectors in $\mathbb{R}^p$ with mean zero, defined on the probability space

$(\Omega, \mathcal{A}, \mathrm{P})$, and let $\check{T} = \max_{1\le j\le p}\sqrt{n}\mathbb{E}_n[X_{ij}]$. For the sake of simplicity, we assume that there is some constant $D_n \ge 1$ such that

$$|X_{ij}| \le D_n,\ a.s.,\ 1 \le i \le n; 1 \le j \le p.$$

We follow the other notation used in Section 7.1. In addition, define

$$S_l = \sum_{i\in I_l} X_i,\ S'_l = \sum_{i \in J_l} X_i,$$

and let $\{\widetilde{S}_l\}_{l=1}^m$ and $\{\widetilde{S}'_l\}_{l=1}^m$ be two independent sequences of random vectors in $\mathbb{R}^p$ such that

$$\widetilde{S}_l \stackrel{d}{=} S_l,\ \widetilde{S}'_l \stackrel{d}{=} S'_l, 1 \le l \le m.$$

Moreover, let $\check{Y} = (\check{Y}_1, \dots, \check{Y}_p)^T$ be a centered normal random vector with covariance matrix $\mathrm{E}[\check{Y}\check{Y}^T] = (1/(mq))\sum_{l=1}^m \mathrm{E}[S_l S_l^T]$.

Theorem B.1 (High dimensional CLT under dependence)**.** *Suppose that there exist constants* $0 < c_1 \le C_1$ *and* $0 < c_2 < 1/4$ *such that* $c_1 \le \underline{\sigma}^2(q) \le \overline{\sigma}^2(r) \vee \overline{\sigma}^2(q) \le C_1, (r/q)\log^2 p \le C_1 n^{-c_2}$*, and*

$$\max\{qD_n \log^{1/2} p, rD_n \log^{3/2} p, \sqrt{q}D_n \log^{7/2}(pn)\} \le C_1 n^{1/2-c_2}.$$

Then there exist constants $c, C > 0$ *depending only on* c_1, c_2, C_1 *such that*

$$\sup_{t\in\mathbb{R}} |\mathrm{P}(\check{T} \le t) - \mathrm{P}(\max_{1\le j\le p} \check{Y}_j \le t)| \le Cn^{-c} + 2(m-1)b_r.$$

Proof. In this proof, c, C denote generic positive constants depending only on c_1, c_2, C_1; their values may change from place to place. We divide the proof into several steps.

Step 1. (Reduction to independence). We wish to show that

$$\begin{aligned}
&\mathrm{P}\left(\max_{1\le j\le p} \frac{1}{\sqrt{n}} \sum_{l=1}^m \widetilde{S}_{lj} \le t - Cn^{-c}\log^{-1/2} p\right) - n^{-c} - 2(m-1)b_r \\
&\quad \le \mathrm{P}(\check{T} \le t) \\
&\quad \le \mathrm{P}\left(\max_{1\le j\le p} \frac{1}{\sqrt{n}} \sum_{l=1}^m \widetilde{S}_{lj} \le t + Cn^{-c}\log^{-1/2} p\right) + n^{-c} + 2(m-1)b_r.
\end{aligned}$$

We only prove the second inequality; the first inequality follows from the analogous argument. Observe that $\sum_{i=1}^n X_i = \sum_{l=1}^m S_l + \sum_{l=1}^m S'_l + S'_{m+1}$, so that

$$|\max_{1\le j\le p} \sum_{i=1}^n X_{ij} - \max_{1\le j\le p} \sum_{l=1}^m S_{lj}| \le \max_{1\le j\le p} |\sum_{l=1}^m S'_{lj}| + \max_{1\le j\le p} |S'_{m+1,j}|.$$

By Corollary 2.7 in Yu (1994) (see also Eberlein, 1984), we have

$$\sup_{t\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\le j\le p}\sum_{l=1}^{m}S_{lj}\le t\right)-\mathrm{P}\left(\max_{1\le j\le p}\sum_{l=1}^{m}\widetilde{S}_{lj}\le t\right)\right|\le(m-1)b_r,$$
$$\sup_{t>0}\left|\mathrm{P}\left(\max_{1\le j\le p}|\sum_{l=1}^{m}S'_{lj}|>t\right)-\mathrm{P}\left(\max_{1\le j\le p}|\sum_{l=1}^{m}\widetilde{S}'_{lj}|>t\right)\right|\le(m-1)b_q.$$

Hence for every $\delta_1,\delta_2>0$,

$$\begin{aligned}\mathrm{P}(\check{T}\le t)&\le\mathrm{P}\left(\max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{l=1}^{m}\widetilde{S}_{lj}\le t+\delta_1+\delta_2\right)\\&+\mathrm{P}\left(\max_{1\le j\le p}|\frac{1}{\sqrt{n}}\sum_{l=1}^{m}\widetilde{S}'_{lj}|>\delta_1\right)+\mathrm{P}\left(\max_{1\le j\le p}|S'_{m+1,j}|>\sqrt{n}\delta_2\right)+2(m-1)b_r\\&=I+II+III+IV.\end{aligned}$$

Since $|S_{m+1,j}|\le(q+r-1)D_n$ a.s., by taking $\delta_2=2(q+r-1)D_n/\sqrt{n}$ ($\le Cn^{-c}\log^{-1/2}p$), we have $III=0$. Moreover, for every $\epsilon>0$, by Markov's inequality, with $\delta_1=\epsilon^{-1}\mathrm{E}[\max_{1\le j\le p}|n^{-1/2}\sum_{l=1}^{m}\widetilde{S}'_{lj}|]$, $II\le\epsilon$. It remains to bound the magnitude of $\mathrm{E}[\max_{1\le j\le p}|n^{-1/2}\sum_{l=1}^{m}\widetilde{S}'_{lj}|]$. Since $\widetilde{S}'_l, 1\le l\le m$, are independent with $|\widetilde{S}'_{lj}|\le rD_n$ a.s. and $\mathrm{Var}(\widetilde{S}'_{lj})\le r\overline{\sigma}^2(r), 1\le l\le m, 1\le j\le p$, by Lemma A.3, we have

$$\mathrm{E}\left[\max_{1\le j\le p}|\frac{1}{\sqrt{n}}\sum_{l=1}^{m}\widetilde{S}'_{lj}|\right]\le K\left(\sqrt{(r/q)\overline{\sigma}^2(r)\log p}+n^{-1/2}rD_n\log p\right).$$

where K is universal (here we have used the simple fact that $m/n\le 1/q$), so that the left side is bounded by $Cn^{-2c}\log^{-1/2}p$ (by taking c sufficiently small). The conclusion of this step follows from taking $\epsilon=n^{-c}$ so that $\delta_1\le Cn^{-c}\log^{-1/2}p$.

Step 2. (Normal approximation to the sum of independent blocks). We wish to show that

$$\sup_{t\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{l=1}^{m}\widetilde{S}_{lj}\le t\right)-\mathrm{P}\left(\max_{1\le j\le p}\sqrt{(mq)/n}\check{Y}_j\le t\right)\right|\le Cn^{-c}.$$

Since $\widetilde{S}_l, 1\le l\le m$, are independent, we may apply Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2013a) (note that the covariance matrix of $\sqrt{(mq)/n}\check{Y}$ is the same as that of $n^{-1/2}\sum_{l=1}^{m}\widetilde{S}_l$). We wish to verify the conditions of the corollary to this case. Observe that

$$\frac{1}{\sqrt{n}}\sum_{l=1}^{m}\widetilde{S}_{lj}=\frac{1}{\sqrt{m}}\sum_{l=1}^{m}\frac{\widetilde{S}_{lj}}{\sqrt{n/m}},$$

and $\sqrt{q} \le \sqrt{n/m} \le 2\sqrt{q}$ (recall that $q + r \le n/2$). Hence

$$c_1/4 \le \underline{\sigma}^2(q)/4 \le \mathrm{Var}\left(\widetilde{S}_{lj}/\sqrt{n/m}\right) \le \overline{\sigma}^2(q) \le C_1,$$

and $|\widetilde{S}_{lj}/\sqrt{n/m}| \le \sqrt{q}D_n$ a.s., so that the conditions of Corollary 2.1 (i) in Chernozhukov, Chetverikov, and Kato (2013a) are verified with $B_n = \sqrt{q}D_n$, which leads to the assertion of this step (note that $q \le Cn^{1-c}$ so that $m \ge n/(4q) \ge C^{-1}n^c$).

Step 3. (Anti-concentration). We wish to verify that, for every $\epsilon > 0$,

$$\sup_{t\in\mathbb{R}} \mathrm{P}\left(\left|\max_{1\le j\le p} \check{Y}_j - t\right| \le \epsilon\right) \le C\epsilon\sqrt{1 \vee \log(p/\epsilon)}.$$

Indeed, since $\check{Y}$ is a normal random vector with

$$c_1 \le \underline{\sigma}^2(q) \le \mathrm{Var}(\check{Y}_j) \le \overline{\sigma}^2(q) \le C_1, 1 \le \forall j \le p,$$

the desired assertion follows from application of Corollary 1 in Chernozhukov, Chetverikov, and Kato (2013b).

Step 4. (Conclusion). By Steps 1-3, we have

$$\sup_{t\in\mathbb{R}}\left|\mathrm{P}(\check{T} \le t) - \mathrm{P}\left(\max_{1\le j\le p} \sqrt{(mq)/n}\check{Y}_j \le t\right)\right| \le Cn^{-c}.$$

It remains to replace $\sqrt{(mq)/n}$ by 1 on the left side. Observe that

$$1 - \sqrt{(mq)/n} \le 1 - (mq)/n \le 1 - (n/(q+r) - 1)(q/n) = r/(q+r) + q/n,$$

and the right side is bounded by $Cn^{-c}\log^{-1} p$. With this c, by Markov's inequality,

$$\mathrm{P}\left(\left|\max_{1\le j\le p} \check{Y}_j\right| > n^{c/2}\sqrt{\log p}\right) \le Cn^{-c/2},$$

as $\mathrm{E}[|\max_{1\le j\le p} \check{Y}_j|] \le C\sqrt{\log p}$, so that with probability larger than $1 - Cn^{-c/2}$,

$$(1 - \sqrt{(mq)/n})\left|\max_{1\le j\le p} \check{Y}_j\right| \le C'n^{-c/2}\log^{-1/2} p.$$

By using the anti-concentration property of $\max_{1\le j\le p} \check{Y}_j$ (see Step 3), we conclude that

$$\sup_{t\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\le j\le p} \sqrt{(mq)/n}\check{Y}_j \le t\right) - \mathrm{P}\left(\max_{1\le j\le p} \check{Y}_j \le t\right)\right| \le Cn^{-c}.$$

This leads to the conclusion of the theorem. □

An inspection of the proof of the above theorem leads to the following corollary on high dimensional CLT for block sums, where the regularity conditions are weaker than those in Theorem B.1.

Corollary B.1 (High dimensional CLT for block sums)**.** *Suppose that there exist constants $C_1 \geq c_1 > 0$ and $0 < c_2 < 1/2$ such that $c_1 \leq \underline{\sigma}^2(q) \leq \overline{\sigma}^2(q) \leq C_1$, and $\sqrt{q} D_n \log^{7/2}(pn) \leq C_1 n^{1/2-c_2}$. Then there exist constants $c, C > 0$ depending only on c_1, c_2, C_1 such that*

$$\sup_{t\in\mathbb{R}} \left| \mathrm{P}\left(\max_{1\leq j\leq p} \frac{1}{\sqrt{mq}} \sum_{l=1}^m S_{il} \leq t \right) - \mathrm{P}(\max_{1\leq j\leq p} \check{Y}_j \leq t) \right| \leq Cn^{-c} + (m-1)b_r.$$

The following theorem is concerned with validity of the block multiplier bootstrap.

Theorem B.2 (Validity of block multiplier bootstrap)**.** *Let $\epsilon_1, \dots, \epsilon_m$ be independent standard normal random variables, independent of the data X_1^n. Suppose that there exist constants $0 < c_1 \leq C_1$ and $0 < c_2 < 1/2$ such that $c_1 \leq \underline{\sigma}^2(q) \leq \overline{\sigma}^2(q) \leq C_1$ and $q D_n \log^{5/2} p \leq C_1 n^{1/2-c_2}$. Then there exist constants $c, c', C, C' > 0$ depending only on c_1, c_2, C_1 such that, with probability larger than $1 - Cn^{-c} - (m-1)b_r$,*

$$\sup_{t\in\mathbb{R}} \left| \mathrm{P}\left(\max_{1\leq j\leq p} \frac{1}{\sqrt{mq}} \sum_{l=1}^m \epsilon_i S_{il} \leq t \mid X_1^n \right) - \mathrm{P}(\max_{1\leq j\leq p} \check{Y}_j \leq t) \right| \leq C' n^{-c'}. \quad (86)$$

Proof. Here c, c', C, C' denote generic positive constants depending only on c_1, c_2, C_1; their values may change from place to place. By Theorem 2 in Chernozhukov, Chetverikov, and Kato (2013b), the left side on (86) is bounded by $C\widehat{\Delta}^{1/3}\{1 \vee \log(p/\widehat{\Delta})\}^{2/3}$, where

$$\widehat{\Delta} = \max_{1\leq j,k\leq p} |(1/(mq)) \textstyle\sum_{l=1}^m (S_{lj}S_{lk} - \mathrm{E}[S_{lj}S_{lk}])|.$$

Hence it suffices to prove that $\mathrm{P}(\widehat{\Delta} > C' n^{-c'} \log^{-2} p) \leq Cn^{-c} + (m-1)b_r$ with suitable c, c', C, C'. By Corollary 2.7 in Yu (1994), for every $t > 0$,

$$\mathrm{P}(\widehat{\Delta} > t) \leq \mathrm{P}(\widetilde{\Delta} > t) + (m-1)b_r,$$

where $\widetilde{\Delta} = \max_{1\leq j,k\leq p} |(1/(mq)) \sum_{l=1}^m (\widetilde{S}_{lj}\widetilde{S}_{lk} - \mathrm{E}[S_{lj}S_{lk}])|$ (recall that $\widetilde{S}_l, 1 \leq l \leq m$, are independent with $\widetilde{S}_l \stackrel{d}{=} S_l$). Observe that $|\widetilde{S}_{lj}\widetilde{S}_{lk}| \leq q^2 D_n^2$ a.s. and $\mathrm{E}[(\widetilde{S}_{lj}\widetilde{S}_{lk})^2] \leq q^3 D_n^2 \overline{\sigma}^2(q)$. Hence by Lemma A.3, we have

$$\mathrm{E}[\widetilde{\Delta}] \leq C(n^{-1/2} q D_n \sqrt{\log p} + n^{-1} q^2 D_n^2 \log p).$$

Since $q D_n \log^{5/2} p \leq C_1 n^{1/2-c_2}$, the right side is bounded by $C' n^{-c_2} \log^{-2} p$. The conclusion of the theorem follows from application of Markov's inequality. □

References

Allen, R. (2014). Powerful, practical procedures for testing many moment inequalities. Preprint.

Andrews, D.W.K. (2004). The block-block bootstrap: improved asymptotic refinements. *Econometrica* **72** 673-700.

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

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

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

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

Armstrong, T.B. and Chan, H.P. (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.L., 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.

Bickel, P., Ritov, Y., and Tsybakov, A. (2009). Simultaneous analysis of lasso and dantzig selector. *Ann. Statist.* **37** 1705-1732.

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

Bülmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data*. Springer.

Canay, I.A. (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. To appear in *Probab. Theory Related Fields.*

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014a). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787-1818.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014b). Central limit theorems and bootstrap in high dimensions. arXiv:1412.3661.

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. and Rosen, A. (2013). Generalized instrumental variable models. cemmap working paper CWP43/13.

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.

Eberlein, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statist. Probab. Lett.* **2** 291-293.

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

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series.* Springer.

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.

Holmes, T. (2011). The diffusion of Wal-Mart and economies of density. *Econometrica* **79** 253-302.

Ingster, Y.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I-III. *Math. Methods Statist.* **2** 85-114, 171-189, 249-268.

Ingster, Y. I. and Suslina, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models.* Lecture Notes in Statistics **169** Springer.

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

Lahiri, S.N. (2003). *Resampling Methods for Dependent Data.* Springer.

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. (2009). Essays on set estimation and inference with moment inequalities. Thesis, MIT.

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

Paparoditis, E. and Politis, D.N. (2001). Tapered block bootstrap. *Biometrika* **88** 1105-1119.

Paparoditis, E. and Politis, D.N. (2002). The tapered block bootstrap for general statistics from stationary sequences. *Econometrics J.* **5** 131-148.

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

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

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

Romano, J.P. 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.

Ryan, S. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica* **80** 1019-1061.

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

van der Vaart, A.W. and Wellner, J.A. (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.

Wu, W. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci.* **102** 14150-14154.

Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **22** 94-116.

Zhang, X. and Cheng, G. (2014). Bootstrapping high dimensional time series. arXiv:1406.1037.

TABLE 1. Results of Monte Carlo experiments for rejection probability. Equicorrelated data, that is $\mathrm{var}(X_i) = \Sigma$ where $\Sigma_{jk} = 1$ if $j = k$ and $\Sigma_{jk} = \rho$ if $j \neq k$. Design 1: $\mathrm{E}[X_{ij}] = 0$ for all j. Design 2: $\mathrm{E}[X_{ij}] = 0$ for $j \leq 0.1p$ and $\mathrm{E}[X_{ij}] = -0.8$ for $j > 0.1p$.

Design 1: Null Hypothesis is True

Density	p	ρ	test type, with (yes) or without (no) selection					
			SN, no	SN, yes	MB, no	MB, yes	EB, no	EB, yes
$t(4)$	200	0	.047	.044	.065	.064	.056	.055
		0.5	.021	.020	.060	.058	.057	.056
		0.9	.001	.001	.053	.051	.056	.053
	500	0	.042	.042	.061	.059	.048	.046
		0.5	.018	.017	.048	.044	.045	.043
		0.9	.003	.003	.052	.051	.049	.047
	1000	0	.029	.028	.046	.045	.037	.033
		0.5	.018	.018	.065	.065	.062	.059
		0.9	0	0	.055	.055	.052	.050
$U(-\sqrt{3}, \sqrt{3})$	200	0	.048	.047	.058	.056	.063	.062
		0.5	.019	.018	.063	.059	.064	.063
		0.9	.001	.001	.056	.055	.057	.056
	500	0	.041	.037	.055	.054	.054	.053
		0.5	.020	.018	.058	.055	.059	.055
		0.9	.003	.003	.047	.046	.054	.049
	1000	0	.035	.034	.056	.054	.057	.057
		0.5	.021	.021	.057	.055	.062	.058
		0.9	0	0	.054	.053	.056	.054

Design 2: Null Hypothesis is True

Density	p	ρ	test type, with (yes) or without (no) selection					
			SN, no	SN, yes	MB, no	MB, yes	EB, no	EB, yes
$t(4)$	200	0	.006	.049	.007	.060	.006	.058
		0.5	.003	.033	.011	.050	.010	.051
		0.9	.001	.004	.022	.047	.024	.044
	500	0	.003	.048	.006	.057	.005	.051
		0.5	.002	.021	.009	.042	.007	.040
		0.9	.002	.003	.034	.057	.034	.053
	1000	0	.004	.042	.005	.054	.004	.046
		0.5	.003	.023	.015	.041	.012	.043
		0.9	0	.003	.037	.053	.034	.055
$U(-\sqrt{3}, \sqrt{3})$	200	0	.008	.048	.010	.049	.009	.050
		0.5	.002	.028	.014	.047	.011	.048
		0.9	0	.003	.023	.038	.024	.040
	500	0	.006	.049	.009	.056	.009	.056
		0.5	.004	.026	.012	.053	.013	.053
		0.9	0	.005	.026	.044	.026	.046
	1000	0	.005	.047	.005	.058	.006	.053
		0.5	.001	.020	.011	.052	.010	.049
		0.9	0	.003	.033	.059	.035	.059

TABLE 2. Results of Monte Carlo experiments for rejection probability. Autocorrelated data, that is $\mathrm{var}(X_i) = \Sigma$ where $\Sigma_{jk} = \rho^{|j-k|}$. Design 3: $\mathrm{E}[X_{ij}] = 0$ for all j. Design 4: $\mathrm{E}[X_{ij}] = 0$ for $j \le 0.1p$ and $\mathrm{E}[X_{ij}] = -0.8$ for $j > 0.1p$.

Design 3: Null Hypothesis is True

Density	p	ρ	test type, with (yes) or without (no) selection					
			SN, no	SN, yes	MB, no	MB, yes	EB, no	EB, yes
$t(4)$	200	0	.042	.039	.052	.052	.046	.045
		0.5	.037	.034	.049	.047	.044	.041
		0.9	.021	.020	.057	.056	.059	.057
	500	0	.032	.031	.039	.036	.035	.034
		0.5	.044	.041	.058	.057	.052	.051
		0.9	.035	.033	.076	.072	.074	.072
	1000	0	.041	.039	.058	.057	.049	.046
		0.5	.037	.033	.051	.050	.049	.047
		0.9	.033	.031	.065	.062	.064	.060
$U(-\sqrt{3}, \sqrt{3})$	200	0	.039	.038	.055	.049	.057	.055
		0.5	.040	.037	.055	.055	.052	.049
		0.9	.016	.014	.042	.041	.044	.042
	500	0	.042	.042	.053	.050	.056	.052
		0.5	.037	.036	.056	.056	.057	.052
		0.9	.033	.031	.062	.061	.066	.064
	1000	0	.040	.040	.063	.060	.065	.060
		0.5	.054	.051	.076	.072	.080	.077
		0.9	.028	.027	.066	.065	.064	.063

Design 4: Null Hypothesis is True

Density	p	ρ	test type, with (yes) or without (no) selection					
			SN, no	SN, yes	MB, no	MB, yes	EB, no	EB, yes
$t(4)$	200	0	.003	.031	.003	.036	.003	.032
		0.5	.002	.037	.002	.051	.002	.050
		0.9	.002	.020	.006	.049	.005	.046
	500	0	.004	.031	.006	.037	.005	.034
		0.5	.002	.043	.005	.057	.006	.047
		0.9	.003	.016	.003	.041	.003	.045
	1000	0	.004	.035	.006	.046	.004	.035
		0.5	.003	.035	.003	.048	.003	.044
		0.9	.002	.022	.003	.057	.005	.058
$U(-\sqrt{3}, \sqrt{3})$	200	0	.004	.033	.005	.039	.005	.045
		0.5	.006	.042	.006	.051	.006	.055
		0.9	.003	.016	.005	.043	.005	.039
	500	0	.008	.050	.010	.059	.010	.060
		0.5	.004	.046	.006	.053	.005	.052
		0.9	.002	.017	.005	.038	.006	.038
	1000	0	.004	.048	.005	.059	.005	.059
		0.5	.006	.037	.008	.043	.007	.044
		0.9	.002	.022	.007	.039	.006	.039

TABLE 3. Results of Monte Carlo experiments for rejection probability. Equicorrelated data, that is $\mathrm{var}(X_i) = \Sigma$ where $\Sigma_{jk} = 1$ if $j = k$ and $\Sigma_{jk} = \rho$ if $j \neq k$. Design 5: $\mathrm{E}[X_{ij}] = 0.05$ for all j. Design 6: $\mathrm{E}[X_{ij}] = 0.05$ for $j \leq 0.1p$ and $\mathrm{E}[X_{ij}] = -0.75$ for $j > 0.1p$.

Design 5: Null Hypothesis is False

Density	p	ρ	test type, with (yes) or without (no) selection					
			SN, no	SN, yes	MB, no	MB, yes	EB, no	EB, yes
$t(4)$	200	0	.696	.692	.749	.739	.729	.722
		0.5	.244	.242	.403	.388	.390	.384
		0.9	.060	. 058	.284	.279	.286	.278
	500	0	.795	.786	.873	.862	.831	.820
		0.5	.206	.197	.393	.380	.373	.365
		0.9	.042	.041	.320	.312	.314	.301
	1000	0	.816	.808	.887	.879	.843	.837
		0.5	.195	.193	.387	.385	.380	.369
		0.9	.028	.026	.280	.275	.276	.273
$U(-\sqrt{3}, \sqrt{3})$	200	0	.711	.695	.782	.768	.781	.773
		0.5	.231	.228	.374	.368	.380	.372
		0.9	.054	.051	.271	.264	.272	.267
	500	0	.770	.759	.850	.839	.859	.852
		0.5	.215	.212	.388	.384	.395	.387
		0.9	.034	.032	.288	.283	.284	.278
	1000	0	.812	.797	.879	.869	.882	.876
		0.5	.211	.209	.416	.404	.413	.408
		0.9	.023	.022	.287	.285	.285	.281

Design 6: Null Hypothesis is False

Density	p	ρ	test type, with (yes) or without (no) selection					
			SN, no	SN, yes	MB, no	MB, yes	EB, no	EB, yes
$t(4)$	200	0	.095	.480	.118	.509	.106	.503
		0.5	.069	.254	.141	.334	.144	.329
		0.9	.023	.103	.182	.302	.180	.288
	500	0	.138	.584	.169	.637	.152	.607
		0.5	.064	.245	.151	.345	.148	.332
		0.9	.018	.073	.210	.278	.201	.280
	1000	0	.170	.630	.210	.697	.189	.660
		0.5	.050	.210	.149	.361	.146	.339
		0.9	.015	.056	.224	.303	.211	.301
$U(-\sqrt{3}, \sqrt{3})$	200	0	.115	.486	.127	.520	.134	.515
		0.5	.061	.253	.128	.354	.128	.360
		0.9	.021	.095	.181	.266	.183	.261
	500	0	.112	.564	.139	.609	.147	.608
		0.5	.074	.251	.166	.357	.163	.362
		0.9	.021	.070	.187	.274	.183	.274
	1000	0	.149	.607	.180	.672	.189	.660
		0.5	.077	.240	.182	.367	.183	.362
		0.9	.014	.057	.205	.266	.203	.272

TABLE 4. Results of Monte Carlo experiments for rejection probability. Autocorrelated data, that is $\mathrm{var}(X_i) = \Sigma$ where $\Sigma_{jk} = \rho^{|j-k|}$. Design 7: $\mathrm{E}[X_{ij}] = 0.05$ for all j. Design 8: $\mathrm{E}[X_{ij}] = 0.05$ for $j \le 0.1p$ and $\mathrm{E}[X_{ij}] = -0.75$ for $j > 0.1p$.

Design 7: Null Hypothesis is False

Density	p	ρ	test type, with (yes) or without (no) selection					
			SN, no	SN, yes	MB, no	MB, yes	EB, no	EB, yes
$t(4)$	200	0	.728	.717	.791	.782	.769	.755
		0.5	.666	.658	.721	.711	.701	.693
		0.9	.364	.362	.535	.525	.539	.528
	500	0	.765	.755	.826	.818	.795	.787
		0.5	.747	.733	.814	.807	.803	.789
		0.9	.437	.428	.663	.652	.659	.648
	1000	0	.826	.813	.900	.898	.862	.852
		0.5	.793	.785	.873	.866	.850	.840
		0.9	.529	.517	.731	.718	.727	.718
$U(-\sqrt{3}, \sqrt{3})$	200	0	.695	.680	.756	.748	.763	.751
		0.5	.658	.644	.733	.725	.735	.725
		0.9	.356	.348	.544	.532	.536	.533
	500	0	.751	.740	.825	.814	.829	.820
		0.5	.737	.730	.806	.799	.816	.805
		0.9	.464	.455	.662	.655	.651	.644
	1000	0	.789	.779	.886	.880	.885	.876
		0.5	.763	.752	.855	.846	.856	.843
		0.9	.503	.498	.734	.721	.729	.719

Design 8: Null Hypothesis is False

Density	p	ρ	test type, with (yes) or without (no) selection					
			SN, no	SN, yes	MB, no	MB, yes	EB, no	EB, yes
$t(4)$	200	0	.103	.545	.128	.577	.115	.567
		0.5	.106	.409	.122	.460	.124	.455
		0.9	.051	.191	.087	.341	.088	.348
	500	0	.125	.555	.154	.605	.143	.573
		0.5	.129	.519	.167	.575	.157	.558
		0.9	.060	.234	.105	.396	.104	.390
	1000	0	.148	.644	.201	.685	.169	.663
		0.5	.137	.584	.168	.656	.163	.627
		0.9	.085	.298	.141	.474	.141	.466
$U(-\sqrt{3}, \sqrt{3})$	200	0	.113	.466	.127	.501	.131	.510
		0.5	.106	.414	.128	.465	.126	.461
		0.9	.054	.189	.092	.312	.093	.323
	500	0	.158	.581	.192	.628	.185	.632
		0.5	.125	.494	.146	.552	.148	.558
		0.9	.050	.237	.103	.375	.100	.378
	1000	0	.148	.562	.188	.630	.189	.637
		0.5	.149	.577	.195	.641	.196	.633
		0.9	.074	.282	.130	.445	.132	.449

Supplement to "Testing many moment inequalities"

by V. Chernozhukov, D. Chetverikov, and K. Kato

MIT, UCLA, and University of Tokyo

Appendix D. Details on equations (10) and (11) in the main text

In this section, we continue discussion of the "Dynamic model of imperfect competition" example presented in Section 2. In particular, we explain how the jackknife procedure leads to equations (10) and (11), which are needed for inference in that example. We continue to assume that the data consist of observations on n i.i.d. markets.

The validity of the jackknife procedure and equations (10) and (11) relies upon the following linear expansions:

$$\begin{aligned}&\sqrt{n}(\widehat{V}_j(s,\sigma_j(\widehat{\alpha}_n),\sigma_{-j}(\widehat{\alpha}_n),\theta)-V_j(s,\sigma_j,\sigma_{-j},\theta))\\&\quad=\frac{1}{\sqrt{n}}\sum_{k=1}^n\psi_{kj}(s,\sigma_j,\sigma_{-j},\theta)+O_P\left(\frac{1}{n}\right)\end{aligned}$$

and

$$\begin{aligned}&\sqrt{n}(\widehat{V}_j(s,\sigma'_j,\sigma_{-j}(\widehat{\alpha}_n),\theta)-V_j(s,\sigma'_j,\sigma_{-j},\theta))\\&\quad=\frac{1}{\sqrt{n}}\sum_{k=1}^n\psi'_{kj}(s,\sigma'_j,\sigma_{-j},\theta)+O_P\left(\frac{1}{n}\right)\end{aligned}$$

where ψ_{kj} and ψ'_{kj} are influence functions depending only on the data for the market k and satisfying

$$\mathrm{E}[\psi_{ij}(s,\sigma_j,\sigma_{-j},\theta)]=0 \text{ and } \mathrm{E}[\psi'_{ij}(s,\sigma'_j,\sigma_{-j},\theta)]=0. \tag{87}$$

These are standard expansions that hold in many settings, so for brevity, we do not discuss the regularity conditions behind them. Further, considering leave-market-i-out estimates $\widehat{\alpha}_n^{-i}$, $\widehat{V}_j^{-i}(s,\sigma_j(\widehat{\alpha}_n^{-i}),\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta)$, and $\widehat{V}_j^{-i}(s,\sigma'_j,\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta)$ from the main text, we obtain

$$\begin{aligned}&\sqrt{n-1}(\widehat{V}_j^{-i}(s,\sigma_j(\widehat{\alpha}_n^{-i}),\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta)-V_j(s,\sigma_j,\sigma_{-j},\theta))\\&\quad=\frac{1}{\sqrt{n-1}}\sum_{k=1;\,k\neq i}^n\psi_{kj}(s,\sigma_j,\sigma_{-j},\theta)+O_P\left(\frac{1}{n-1}\right)\end{aligned}$$

and

$$\sqrt{n-1}(\widehat{V}_j^{-i}(s,\sigma_j',\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta)-V_j(s,\sigma_j',\sigma_{-j},\theta))$$
$$=\frac{1}{\sqrt{n-1}}\sum_{k=1;\,k\neq i}^{n}\psi_{kj}'(s,\sigma_j',\sigma_{-j},\theta)+O_P\left(\frac{1}{n-1}\right).$$

Hence, we have for all $i=1,\dots,n$,

$$\begin{aligned}\widetilde{X}_{ij}(s,\theta) &:= n\widehat{V}_j(s,\sigma_j(\widehat{\alpha}_n),\sigma_{-j}(\widehat{\alpha}_n),\theta)\\ &\quad -(n-1)\widehat{V}_j^{-i}(s,\sigma_j(\widehat{\alpha}_n^{-i}),\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta)\\ &= V_j(s,\sigma_j,\sigma_{-j},\theta)+\psi_{ij}(s,\sigma_j,\sigma_{-j},\theta)+O_P(n^{-1/2})\end{aligned}$$

and

$$\begin{aligned}\widetilde{X}_{ij}'(s,\sigma_j',\theta) &:= n\widehat{V}_j(s,\sigma_j',\sigma_{-j}(\widehat{\alpha}_n),\theta)\\ &\quad -(n-1)\widehat{V}_j^{-i}(s,\sigma_j',\sigma_{-j}(\widehat{\alpha}_n^{-i}),\theta)\\ &= V_j(s,\sigma_j',\sigma_{-j},\theta)+\psi_{ij}'(s,\sigma_j',\sigma_{-j},\theta)+O_P(n^{-1/2}).\end{aligned}$$

Conclude that

$$\widehat{X}_{ij}(s,\sigma_j',\theta):=\widetilde{X}_{ij}'(s,\sigma_j',\theta)-\widetilde{X}_{ij}(s,\theta)=X_{ij}(s,\sigma_j',\theta)+O_p(n^{-1/2})$$

where

$$\begin{aligned}X_{ij}(s,\sigma_j',\theta) :=& V_j(s,\sigma_j',\sigma_{-j},\theta)-V_j(s,\sigma_j,\sigma_{-j},\theta)\\ &+\psi_{ij}'(s,\sigma_j',\sigma_{-j},\theta)-\psi_{ij}(s,\sigma_j,\sigma_{-j},\theta).\end{aligned}$$

Combining these equalities with (87) implies (10) and (11) from the main text and completes the derivation.

(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