INFERENCE ON CAUSAL AND STRUCTURAL PARAMETERS USING 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 is a variety of economic applications where solving this problem allows to carry out inference on causal and structural parameters; a notable example is the market structure model of Ciliberto and Tamer (2009) where $p = 2^{m+1}$ with m being the number of firms that could possibly enter the market. We consider the test statistic given by the maximum of p Studentized (or t-type) inequality-specific 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 the selection of uninformative inequalities that are far from being binding and a 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 the 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 the correlation structure between p Studentized statistics, and also hold uniformly with respect to suitably large classes of underlying distributions. Moreover, in the online supplement, we show validity of a test based on the block multiplier bootstrap in the case of dependent data under some general mixing conditions.

1. Introduction

In recent years, the moment inequalities framework has developed into a powerful tool for inference on causal and structural parameters in partially identified models. Many papers studied models with a finite and fixed (and

Date: First public version: December 2013 (arXiv:1312.7614v1). This version: October 19, 2018.

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. This paper was previously circulated under the title "Testing many moment inequalities."

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 is a 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 the 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 the 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 (2013), and Chesher and Rosen (2013) where moment inequalities are used to provide sharp identification regions for parameters in partially identified models. In all these applications, testing moment inequalities allows

[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.

to carry out inference on structural and causal parameters. In addition, we note that, as explained in Shah and Peters (2018), our results help to test conditional independence, a concept that plays a particularly important role in causal machine learning; see Pearl (2009).

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. 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.[2]

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 the market structure in the US airline industry,[3] Holmes (2011) who estimated the dynamic model of the 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 a certain continuity and tightness structure, which the literature on conditional moment inequalities heavily exploits/relies upon. Our approach works even if such structure is not available and can handle many unstructured moment inequalities. In addition, when such structure is available, our bootstrap methods automatically exploit it leading to powerful tests of structured moment inequalities arising from conversion of a small or large number of conditional moment 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, which is already a large number. 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.

[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.

$X_i = (X_{i1}, \dots, X_{ip})^T$, with a common distribution denoted by $\mathcal{L}_X$. 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) inequality-specific 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 Appendix A of the online supplement. Bootstrap methods are simulation-based and computationally harder. However, an important feature of bootstrap methods is that they take into account the 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 the one-step methods, that is, methods without the inequality selection procedures.

Moreover, we develop novel three-step methods by incorporating double inequality selection procedures. The three-step methods are suitable in parametric models defined via moment inequalities and allow to drop *weakly* *informative* inequalities in addition to uninformative inequalities.[6] Specifically, consider the model consisting of inequalities $\mathrm{E}[g_j(\xi,\theta)] \leq 0$ for all $j = 1, \dots, p$ where ξ is a vector of observable random variables, θ a vector of structural or causal parameters, and $g_1, \dots, g_p$ a set of known 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 $\xi_1, \dots, \xi_n$, 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 the violation of the null hypothesis when θ is close to θ_0.

We prove validity of these methods for computing the critical values, uniformly in suitable classes of distributions $\mathcal{L}_X$. We derive non-asymptotic bounds on the rejection probabilities, where "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.[7]

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, as a function of the number of moment inequalities p, our critical values are always bounded from above by a slowly varying $(\log p)^{1/2}$ (up to a multiplicative constant). This implies that instead of making a subjective choice of inequalities, the researcher should use all (or at least a large set of) available inequalities since using more inequalities gives much larger values of the test statistic when added inequalities violate H_0. This feature of our methods is akin to that in modern high-dimensional/big-data techniques like the Lasso and the Dantzig selector that allow for the variable selection in exchange for small cost in the precision of model estimates; see, for example, Bickel, Ritov, and Tsybakov (2009) for an analysis and discussion of the methods of estimating high-dimensional models.

Our results can also be used for the construction of confidence regions for identifiable parameters in partially identified models defined by moment inequalities. In particular, we show in Appendix A of the online supplement how to use our results for constructing confidence regions that are *asymptotically honest*, with the coverage being correct uniformly in suitably large classes of underlying distributions.

[6] 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.

[7] When p is small relative to n, other tests, e.g. the quasi likelihood-ratio test may be more powerful than the methods developed here; see Section 3 for further discussion.

Moreover, we consider two extensions of our results in Appendix B of the online supplement. In the first extension, we consider testing many moment inequalities for dependent data. 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 literature on testing (unconditional) moment inequalities is large; see White (2000), 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 (2014). 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 (2015), Chetverikov (2017), and Armstrong and Chan (2016). 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 (2017); and he considered moment re-centering procedure similar to that developed in Romano, Shaikh, and Wolf (2014). The latter two possibilities were already noted in the previous versions of our paper.[8]

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 state results on their validity. In Section 5, we discuss power properties of our methods. In Section 6, we describe Monte Carlo simulations shedding light on how our methods perform in finite samples. Additional results, as well as all the proofs and the results of Monte Carlo simulations, are provided in the online supplement.

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

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 $p \geq 2$.

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.[9] 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).[10] 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 the (exogenous) characteristics of the market as well as the characteristics of the firms that are observed and not observed by the

[9] 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 $\mathcal{I}_n$ grows sufficiently fast with n, the test of (3) based on (4) will be consistent.

[10] The market structure model is also often referred to as an entry game.

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 the 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 *an* equilibrium 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 $D = d$ 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

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

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. Finally, one can produce even larger set of inequalities in this example using the bounds of Galichon and Henry (2011); see Section 6.3 for details. 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}'s are 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 Bellman 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

[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 Appendix B.1 of the online supplement. 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.

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 $\alpha = (\alpha_1, \alpha_2)$, that is $P(S_{t+1}|S_t, A_t) = P(S_{t+1}|S_t, A_t, \alpha_1)$ and $\sigma_j(s, \nu_j) = \sigma_j(s, \nu_j, \alpha_2)$, and that the first stage yields a $\sqrt{n}$-consistent estimator $\widehat{\alpha}_n = (\widehat{\alpha}_{n,1}, \widehat{\alpha}_{n,2})$ of $\alpha = (\alpha_1, \alpha_2)$.[12] Using $\widehat{\alpha}_{n,1}$, one can estimate the transition probability function by $P(S_{t+1}|S_t, A_t, \widehat{\alpha}_{n,1})$, 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). 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 function $P(S_{t+1}|S_t, A_t, \alpha)$. Then, on the second stage, one can test the 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

$$\widehat{V}_j(s, \widehat{\sigma}_j, \widehat{\sigma}_{-j}, \theta) \geq \widehat{V}_j(s, \sigma'_j, \widehat{\sigma}_{-j}, \theta) \tag{9}$$

where $\widehat{\sigma}_j = \sigma_j(\widehat{\alpha}_{n,2})$ and $\widehat{\sigma}_{-j} = \sigma_{-j}(\widehat{\alpha}_{n,2})$ are the estimated policy functions for the firm j and all of its rivals, respectively. Inequalities (9) can be used to test hypotheses about the 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. To explain the procedure, assume that the data consist of observations on n i.i.d markets. Let $\widehat{V}_j^{-i}(s, \sigma', \theta)$ and $\widehat{\sigma}^{-i}$ denote the leave-market-i-out estimates of $V_j(s, \sigma', \theta)$ and σ, respectively. Define

$$\widetilde{X}_{ij}(s, \theta) := n\widehat{V}_j(s, \widehat{\sigma}_j, \widehat{\sigma}_{-j}, \theta) - (n-1)\widehat{V}_j^{-i}(s, \widehat{\sigma}_j^{-i}, \widehat{\sigma}_{-j}^{-i}, \theta)$$

and

$$\widetilde{X}'_{ij}(s, \sigma'_j, \theta) := n\widehat{V}_j(s, \sigma'_j, \widehat{\sigma}_{-j}, \theta) - (n-1)\widehat{V}_j^{-i}(s, \sigma'_j, \widehat{\sigma}_{-j}^{-i}, \theta).$$

[12] Estimation of α_1 is simple; for example, it can be estimated by the maximum likelihood method. Estimation of α_2 is more complicated since the functions $\sigma_j(s, \nu_j)$ depend on unobservable ν_j's and requires additional assumptions. When the set $\mathcal{A}$ is finite, for example, one can assume that the shock ν_j is additively separable in the profit function, so that $\pi_j(A_t, S_t, \nu_{jt}) = \widetilde{\pi}_j(A_t, S_t) + \nu_i(A_{jt})$, where the vector $\{\nu_i(A)\}_{A\in\mathcal{A}}$ consists of i.i.d. random variables, and use the methods of Hotz and Miller (1993) to estimate α_2; see Bajari, Benkard, and Levin (2007) for details.

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

$$\widehat{X}_{ij}(s,\sigma_j',\theta) = X_{ij}(s,\sigma_j',\theta) + o_P(1) \tag{10}$$

for some $X_{ij}(s,\sigma_j',\theta)$ satisfying

$$\mathrm{E}[X_{ij}(s,\sigma_j',\theta)] = V_j(s,\sigma_j',\sigma_{-j},\theta) - V_j(s,\sigma,\theta) \le 0, \tag{11}$$

where $X_{ij}(s,\sigma_j',\theta)$'s are independent across markets $i = 1,\dots,n$. We provide some details on the derivation of (10) and (11) in Appendix C of the online supplement. Now we can use the results of Appendix B.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, in addition, we use $(\widehat{V}_j(s,\sigma_j',\widehat{\sigma}_{-j},\theta) - \widehat{V}_j(s,\widehat{\sigma}_j,\widehat{\sigma}_{-j},\theta))$ instead of $\widehat{\mu}_j = n^{-1}\sum_{i=1}^n \widehat{X}_{ij}(s,\sigma_j',\theta)$ in the numerator of our test statistic defined in (13).[13] Thus, this example fits into our framework as well.[14]

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

[13] Note that one of the conditions of Theorem B.2 is that (58) holds with $\widehat{\mu}_{j,0} = n^{-1}\sum_{i=1} X_{ij}(s,\sigma_j',\theta)$ in our case, and since we can only guarantee that $\widehat{X}_{ij}(s,\sigma_j',\theta) - X_{ij}(s,\sigma_j',\theta) = O_P(n^{-1/2})$ as in (10), this condition may not be satisfied if we define $\widehat{\mu}_j = n^{-1}\sum_{i=1}^n \widehat{X}_{ij}(s,\sigma_j',\theta)$. This condition is satisfied, however, under mild regularity conditions, if we define $\widehat{\mu}_j = \widehat{V}_j(s,\sigma_j',\widehat{\sigma}_{-j},\theta) - \widehat{V}_j(s,\widehat{\sigma}_j,\widehat{\sigma}_{-j},\theta)$; see the online supplement for details.

[14] 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.

is natural to consider statistics that take large values when some of $\widehat{\mu}_j$'s are large. In this paper, we focus on the statistic that takes large values when at least one of $\widehat{\mu}_j$'s 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 the power evenly among the different moment inequalities $\mu_j \leq 0$. 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.

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

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

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

$$T' = \sum_{j=1}^{p} \left(\max\{\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j, 0\}\right)^2. \tag{15}$$

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 quasi likelihood-ratio test statistic of the form

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

where $\widehat{\mu} = (\widehat{\mu}_1, \dots, \widehat{\mu}_p)^T$, $t = (t_1, \dots, t_p)^T \leq 0$ means $t_j \leq 0$ 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. On the other hand, when p is small relative to n, the test statistic T'' may lead to more powerful tests than those based on T and T' since it takes into account the correlation structure between the inequalities, like GMM does in the setting of moment equalities. For the rest of the paper, we focus on the statistic T and do not provide critical values for the tests based on T' and T''.

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 asymptotically. The methods are essentially ordered by increasing computational complexity, increasing strength of required conditions, but also increasing power. We note, however, that all our methods require only mild conditions on the underlying distributions and are computationally rather simple.

The basic idea for construction of critical values for T lies in the fact that under H_0,

$$T \leq \max_{1 \leq j \leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j, \tag{16}$$

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 $\max_{1 \leq j \leq p} \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j$. We consider two approaches to construct such critical values: self-normalized and 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

$$Z_{ij} = (X_{ij} - \mu_j)/\sigma_j, \text{ and } Z_i = (Z_{i1}, \dots, Z_{ip})^T. \tag{17}$$

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

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

($M_{n,k}$ and B_n depend on n since $p = p_n$ (implicitly) depends on n.) Note that by Jensen's inequality, $B_n \ge M_{n,4} \ge M_{n,3} \ge 1$. In addition, if Z_{ij}'s are all bounded by a constant C almost surely, we have $C \ge B_n$. These inequalities are useful to get a sense of various conditions on $M_{n,3}$, $M_{n,4}$, and B_n imposed in the theorems below.

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 (16), under H_0,

$$\mathrm{P}(T > c) \le \sum_{j=1}^{p} \mathrm{P}(\sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j > c). \tag{18}$$

At a first sight, this bound might look too crude when p is large since, as long as X_{ij}'s have polynomial tails, applying, for example, the Markov inequality would only allow us to show that the right-hand side of (18) is bounded from above by α when c is growing *polynomially* fast with p, and using such c would yield a test with low power. However, the Markov inequality is far from being sharp here. Instead, we will exploit the self-normalizing nature of the quantity $\sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j$ to show that the right-hand side of (18) is bounded from above by α, up to a vanishing term, even if c is growing *logarithmically* fast with p. Using such c will in turn yield a test with much better power properties.

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 $U_j \ge 0$. Hence under H_0,

$$\mathrm{P}(T > c) \le \sum_{j=1}^{p} \mathrm{P}\left(U_j > c/\sqrt{1 + c^2/n} \right), \ c \ge 0. \tag{19}$$

Now, the moderate deviation inequality for self-normalized sums of Jing, Shao, and Wang (2003) (see Lemma D.1 in the online supplement) 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

$$c^{SN}(\alpha) = \frac{\Phi^{-1}(1-\alpha/p)}{\sqrt{1-\Phi^{-1}(1-\alpha/p)^2/n}}, \tag{20}$$

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,*

$$\mathrm{P}(T > c^{SN}(\alpha)) \leq \alpha\left[1 + Kn^{-1/2}M_{n,3}^3\{1+\Phi^{-1}(1-\alpha/p)\}^3\right], \tag{21}$$

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

$$M_{n,3}^3 \log^{3/2}(p/\alpha) \leq C_1 n^{1/2-c_1}, \tag{22}$$

then there exists a positive constant C depending only on C_1 such that under H_0,

$$\mathrm{P}(T > c^{SN}(\alpha)) \leq \alpha + Cn^{-c_1}. \tag{23}$$

Moreover, this bound holds uniformly over all distributions $\mathcal{L}_X$ satisfying (12) and (22). In addition, if (22) holds, all components of X_1 are independent, $\mu_j = 0$ for all $1 \leq j \leq p$, and $p = p_n \to \infty$, then

$$\mathrm{P}(T > c^{SN}(\alpha)) \to 1 - e^{-\alpha}. \tag{24}$$

Comment 4.1 (On conditions of Theorem 4.1)**.** Since condition (22) is abstract, it is instructive to see how this condition looks in particular examples. Suppose, for example, that all X_{ij}'s are Gaussian. Then all Z_{ij}'s are standard Gaussian, and so $\mathrm{E}[|Z_{1j}|^3] = (8/\pi)^{1/2}$. Hence, it follows that $M_{n,3} = \max_{1\leq j\leq p}(\mathrm{E}[|Z_{1j}|^3])^{1/3} = (8/\pi)^{1/6}$, and condition (22) reduces to

$\log^{3/2}(p/\alpha) \leq C_1 n^{1/2-c_1}$ (with a different constant C_1). When α is independent of n, the condition further reduces to $(\log^3 p)/n \leq C_1 n^{-c_1}$ (with possibly different constants c_1 and C_1).

Comment 4.2 (Relaxing conditions of Theorem 4.1)**.** The theorem assumes that $\max_{1\leq j\leq p} \mathrm{E}[|X_{1j}|^3] < \infty$ (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 $\max_{1\leq j\leq p}(\mathrm{E}[|Z_{1j}|^{2+\nu}])^{1/(2+\nu)}$ for $0 < \nu \leq 1$, which would further weaken moment conditions; however, for the sake of simplicity of presentation, we do not explore this generalization.

Comment 4.3 (On conservativeness of the one-step SN method)**.** The last asserted claim of Theorem 4.1, (24), shows that when p is large, all components of X_1 are independent, and all inequalities satisfy the null and are binding, the one-step SN method is approximately non-conservative. Indeed, the nominal level α is typically small, e.g. 5% or 10%, so that $e^{-\alpha} \approx 1 - \alpha$, and the probability of rejecting the null is approximately α in this case.

Comment 4.4 (Comparison with the classical Bonferroni procedure)**.** The classical Bonferroni approach to test (1) against (2) would be to compare the statistic T with the Bonferroni critical value $c^{Bon}(\alpha) = \Phi^{-1}(1 - \alpha/p)$. It is straightforward to show using standard techniques that this approach works (controls size) when p is much smaller than n or X_i's are Gaussian. In contrast, our techniques do not require these conditions, which is important because it allows us to test many moment inequalities in a wide variety of settings, without assuming Gaussianity. In addition, using our techniques, it is possible to show that the Bonferroni approach also works under the same conditions as those required for our SN method; see Theorem D.1 in the online supplement.

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 (16) 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 $0 < \beta_n < \alpha/2$ be some constant. For generality, we allow β_n to depend on n; in particular, β_n is allowed to decrease to zero as the sample size n increases. Let $c^{SN}(\beta_n)$ be the SN critical value with size β_n, and define the set $\widehat{J}_{SN} \subset \{1, \dots, p\}$ by

$$\widehat{J}_{SN} := \left\{ j \in \{1, \dots, p\} : \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j > -2c^{SN}(\beta_n) \right\}. \tag{25}$$

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

$$c^{SN,2S}(\alpha) = \begin{cases} \frac{\Phi^{-1}(1-(\alpha-2\beta_n)/\widehat{k})}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta_n)/\widehat{k})^2/n}}, & \text{if } \widehat{k} \geq 1, \\ 0, & \text{if } \widehat{k} = 0. \end{cases} \tag{26}$$

The following theorem establishes validity of this critical value.

Theorem 4.2 (Validity of two-step SN method)**.** *Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$\begin{aligned} M_{n,3}^3 \log^{3/2}\left(\frac{p}{\beta_n \wedge (\alpha - 2\beta_n)}\right) &\leq C_1 n^{1/2-c_1}, \\ \text{and } \ B_n^2 \log^2(p/\beta_n) &\leq C_1 n^{1/2-c_1}. \end{aligned} \tag{27}$$

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

$$\mathrm{P}(T > c^{SN,2S}(\alpha)) \leq \alpha + Cn^{-c}. \tag{28}$$

Moreover, this bound holds uniformly over all distributions $\mathcal{L}_X$ satisfying (12) and (27). In addition, if all components of X_1 are independent, $\mu_j = 0$ for all $1 \leq j \leq p$, $p = p_n \to \infty$, and $\beta_n \to 0$, then

$$\mathrm{P}(T > c^{SN,2S}(\alpha)) \to 1 - e^{-\alpha}. \tag{29}$$

Comment 4.5 (Comparing conditions of one-step and two-step SN methods)**.** Observe that the condition (27) required for the validity of the two-step SN method in Theorem 4.2 is stronger than the condition (22) required for the validity of the one-step SN method in Theorem 4.1. To see the meaning of (27) under primitive conditions, suppose that all X_{ij}'s are Gaussian. Then all Z_{ij}'s are standard Gaussian, and so $B_n = (\mathrm{E}[\max_{1\leq j\leq p} Z_{1j}^4])^{1/4} \leq C(\log p)^{1/2}$ for some constant $C > 0$. Hence, given that $M_{n,3} \leq C$ in this case and $\beta_n < 1$, it follows that condition (27) is implied by $\log^3(p/(\beta_n \wedge (\alpha - 2\beta_n))) \leq C_1 n^{1/2-c_1}$ (with a different constant C_1). Hence, if $cn^{-1/C} \leq \beta_n \leq \alpha/2 - c$, it follows that condition (27) holds when $\log^6 p/n \leq C_1 n^{-c_1}$ (with different constants c_1 and C_1).

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 (17). *When p is fixed*, the central limit theorem guarantees that as $n\to\infty$,

$$\sqrt{n}\mathbb{E}_n[Z_i]\overset{d}{\to}Y,\text{ with }Y=(Y_1,\dots,Y_p)^T\sim N(0,\mathrm{E}[Z_1Z_1^T]),$$

which, by the continuous mapping theorem, implies that

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

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, 2013, 2017) 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.*[15] This result implies that we can still use the $(1-\alpha)$-quantile of the distribution of $\max_{1\leq j\leq p}Y_j$ even when p grows with n and is potentially much larger than n.[16]

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

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

[16] Some applications of this result can be found in Chetverikov (2017, 2012), Wasserman, Kolar and Rinaldo (2013), and Chazal, Fasy, Lecci, Rinaldo, and Wasserman (2013).

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
$$W^{MB} = \max_{1 \leq j \leq p} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \widehat{\mu}_j)]}{\widehat{\sigma}_j}. \tag{30}$$

3. Calculate $c^{MB}(\alpha)$ as
$$c^{MB}(\alpha) = \text{conditional } (1-\alpha)\text{-quantile of } W^{MB} \text{ given } X_1^n. \tag{31}$$

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
$$W^{EB} = \max_{1 \leq j \leq p} \frac{\sqrt{n}\mathbb{E}_n[X_{ij}^* - \widehat{\mu}_j]}{\widehat{\sigma}_j}. \tag{32}$$

3. Calculate $c^{EB}(\alpha)$ as
$$c^{EB}(\alpha) = \text{conditional } (1-\alpha)\text{-quantile of } W^{EB} \text{ given } X_1^n. \tag{33}$$

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
$$\left(\frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \widehat{\mu}_j)]}{\widehat{\sigma}_j} \right)_{1 \leq j \leq p}$$
has the centered normal distribution with covariance matrix
$$\mathbb{E}_n \left[\frac{(X_{ij} - \widehat{\mu}_j)}{\widehat{\sigma}_j} \frac{(X_{ik} - \widehat{\mu}_k)}{\widehat{\sigma}_k} \right], \ 1 \leq j, k \leq p, \tag{34}$$
which should be close to the covariance matrix of the vector Y. Indeed, by Theorem 2 in Chernozhukov, Chetverikov, and Kato (2015), the primary factor for the bound on the Kolmogorov distance between the conditional distribution of W and the distribution of $\max_{1 \leq j \leq p} Y_j$ is
$$\max_{1 \leq j,k \leq p} \left| \mathbb{E}_n \left[\frac{(X_{ij} - \widehat{\mu}_j)}{\widehat{\sigma}_j} \frac{(X_{ik} - \widehat{\mu}_k)}{\widehat{\sigma}_k} \right] - \mathrm{E}[Z_{1j} Z_{1k}] \right|,$$
which 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

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

can be well approximated in distibution by the maximum of a random vector with centered normal distribution with covariance matrix (34) 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* $c^B(\alpha)$ *stand 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*

$$(M^3_{n,3} \vee M^2_{n,4} \vee B_n)^2 \log^{7/2}(pn) \leq C_1 n^{1/2-c_1}. \tag{35}$$

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

$$\mathrm{P}(T > c^B(\alpha)) \leq \alpha + Cn^{-c}. \tag{36}$$

In addition, if $\mu_j = 0$ *for all* $1 \leq j \leq p$*, then*

$$|\mathrm{P}(T > c^B(\alpha)) - \alpha| \leq Cn^{-c}. \tag{37}$$

Moreover, both bounds hold uniformly over all distributions $\mathcal{L}_X$ *satisfying (12) and (35).*

Comment 4.6 (High dimension bootstrap CLT)**.** The result (37) 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 are 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 (2013) and Chernozhukov, Chetverikov, and Kato (2017), and may be of interest in many other settings, well beyond the problem of testing many moment inequalities.

Comment 4.7 (Comparison with White, 2000)**.** White (2000) is relevant to our one-step MB/EB methods in the sense that White (2000) considers a max-type statistic for an inequality testing problem and applies bootstrap to calibrate critical values. However, White (2000) does not consider Studentization, and more importantly 1) does not allow the number of inequalities increasing with the sample size, and 2) does not consider inequality selection so that his test would be conservative (see the next subsection on our two-step MB/EB methods). In fact, White (2000) acknowledges the importance of extending his analysis to the case where the number of inequalities increases with the sample size, and explicitly states that "it is natural to consider what happens when l grows with T" [l is the number of inequalities tested and T is the sample size] but " rigorous treatment for our context is beyond our present scope" (White, 2000, p.1110-1111). Our results on the one-step MB/EB methods address this important question in a far more general setting where the number of inequalities can be much larger than the sample size. In addition, our results provided finite sample error bounds that hold uniformly over a wide class of underlying distributions, while White (2000) only derives pointwise asymptotic results on validity of the test.

Comment 4.8 (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.

Comment 4.9 (Comparing conditions of two-step SN method and one-step MB/EB methods)**.** Observe that the condition (35) required for the validity of the one-step MB/EB methods in Theorem 4.3 is stronger than the condition (27) required for the validity of the two-step SN method in Theorem 4.2. To see the meaning of (35) under primitive conditions, suppose that all X_{ij}'s are Gaussian. As in Comment 4.5, it then follows that $M_{n,3} \leq C$ and $B_n \leq C(\log p)^{1/2}$ for some constant C in this case. Moreover, it is easy to see that $M_{n,4} \leq C$ as well. Therefore, condition 4.5 holds if $(\log^9 p)/n \leq C_1 n^{-c_1}$ (with possibly different constants c_1 and C_1).

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 stands 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

$$c^{MB,2S}(\alpha) = \text{conditional } (1-\alpha+2\beta_n)\text{-quantile of } W_{\widehat{J}_{MB}} \text{ given } X_1^n. \quad (38)$$

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

$$c^{EB,2S}(\alpha) = \text{conditional } (1-\alpha+2\beta_n)\text{-quantile of } W_{\widehat{J}_{EB}} \text{ given } X_1^n. \quad (39)$$

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 $c^{B,2S}(\alpha)$ stand either for $c^{MB,2S}(\alpha)$ or $c^{EB,2S}(\alpha)$. Suppose that the assumption of Theorem 4.3 is satisfied. Moreover, suppose that $\log(1/\beta_n) \le C_1 \log n$. Then there exist positive constants c, C depending only on c_1, C_1 such that under H_0,*

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

In addition, if $\mu_j = 0$ for all $1 \le j \le p$, then

$$\mathrm{P}(T > c^{B,2S}(\alpha)) \ge \alpha - 3\beta_n - Cn^{-c},$$

so that under an extra assumption that $\beta_n \le C_1 n^{-c_1}$, then

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

Moreover, all these bounds hold uniformly over all distributions $\mathcal{L}_X$ satisfying (12) and (35).

Comment 4.10. The selection procedure used in the theorem above is most closely related to those in Chernozhukov, Lee, and Rosen (2013) and in Chetverikov (2017). Other selection procedures were suggested in the literature in the framework when p is fixed. Specifically, Romano, Shaikh, and Wolf (2014) 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 (37), we have that $\mu_j \le \widehat{\mu}_j + \widehat{\sigma}_j c^{MB}(\beta_n)/\sqrt{n}$ for all $1 \le j \le p$ with probability $1 - \beta_n$ asymptotically. Therefore, we can replace (16) with the following probabilistic inequality: under H_0,

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

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 (25). 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

$$c^{MB,H}(\alpha) = \text{conditional } (1 - \alpha + 2\beta_n)\text{-quantile of } W_{\widehat{J}_{SN}} \text{ given } X_1^n. \quad (40)$$

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 $c^{B,H}(\alpha)$ stand 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 (35) is verified. Moreover, suppose that $\log(1/\beta_n) \leq C_1 \log n$. 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

$$\mathrm{E}[g_j(\xi, \theta)] \leq 0 \quad \text{for all } j = 1, \dots, p, \quad (41)$$

where ξ is a vector of random variables from a distribution denoted by $\mathcal{L}_\xi$, $\theta = (\theta_1, \dots, \theta_r)^T$ is a vector of parameters in $\mathbb{R}^r$, 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 (41) and that can substantially 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 (41), such inequalities can only provide a weak signal of violation of the hypothesis $\theta = \theta_0$ in the sense that they have $\mu_j \approx 0$, and so it is useful to drop them. 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. Also, we only consider the case when the functions $\theta \mapsto g_j(\xi, \theta)$ are almost surely continuously differentiable, and leave the extension to non-differentiable functions to future work.

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

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

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. Define

$$\begin{aligned} m_j(\xi, \theta) &:= (m_{j1}(\xi, \theta), \dots, m_{jr}(\xi, \theta))^T \\ &:= (\partial g_j(\xi, \theta)/\partial \theta_1, \dots, \partial g_j(\xi, \theta)/\partial \theta_r)^T \end{aligned}$$

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

$$\mathrm{E}[X_{1j}^2] < \infty,\ \sigma_j > 0,\ j = 1, \dots, p, \tag{42}$$

$$\mathrm{E}[V_{1jl}^2] < \infty,\ \sigma^V_{jl} > 0,\ j = 1, \dots, p,\ l = 1, \dots, r. \tag{43}$$

In addition, let

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

be estimators of μ_j and σ_j, respectively, and let

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

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
$$W_{MB}^V = \max_{j,l} \frac{\sqrt{n}|\mathbb{E}_n[\epsilon_i(V_{ijl} - \widehat{\mu}_{jl}^V)]|}{\widehat{\sigma}_{jl}^V}. \tag{44}$$

3. For $\gamma \in (0,1)$, calculate $c^{MB,V}(\gamma)$ as
$$c^{MB,V}(\gamma) = \text{conditional } (1-\gamma)\text{-quantile of } W_{MB}^V \text{ given } \xi_1^n. \tag{45}$$

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
$$W_{EB}^V = \max_{j,l} \frac{\sqrt{n}|\mathbb{E}_n[V_{ijl}^* - \widehat{\mu}_{jl}^V]|}{\widehat{\sigma}_{jl}^V}. \tag{46}$$

3. For $\gamma \in (0,1)$, calculate $c^{EB,V}(\gamma)$ as
$$c^{EB,V}(\gamma) = \text{conditional } (1-\gamma)\text{-quantile of } W_{EB}^V \text{ given } \xi_1^n. \tag{47}$$

For some strictly positive constants c_2 and C_2, let φ_n be a sequence of constants satisfying $\varphi_n \log n \geq c_2$, and let β_n be a sequence of constants satisfying $0 < \beta_n < \alpha/4$ 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:

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

where B stands 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 $4\beta_n$:

$$c^{B,2S}(\alpha) = \text{conditional } (1-\alpha+4\beta_n)\text{-quantile of } W_{\widehat{J}_B \cap \widehat{J}''_B} \text{ given } X_1^n,$$

where $W_{\widehat{J}_B \cap \widehat{J}''_B}$ is either the multiplier or the bootstrap test statistic depending on whether $B = MB$ or EB. The test rejects H_0 if $T^B > c^{B,3S}(\alpha)$.[17]

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

$$Z_{ijl}^V := (V_{ijl} - \mu_{jl}^V)/\sigma_{jl}^V.$$

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

$$M_{n,k}^V := \max_{j,l} \left(\mathrm{E}[|Z_{1jl}^V|^k]\right)^{1/k}, \ k = 3, 4, \ B_n^V := \left(\mathrm{E}\Big[\max_{j,l} (Z_{1jl}^V)^4\Big]\right)^{1/4}.$$

We have the following theorem:

Theorem 4.6 (Validity of three-step MB and EB methods)**.** *Let T^B and $c^{B,3S}(\alpha)$ stand 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*

$$\left(M_{n,3}^3 \vee M_{n,4}^2 \vee B_n\right)^2 \log^{7/2}(pn) \le C_1 n^{1/2-c_1} \tag{48}$$

and

$$\left((M_{n,3}^V)^3 \vee (M_{n,4}^V)^2 \vee B_n^V\right)^2 \log^{7/2}(prn) \le C_1 n^{1/2-c_1}. \tag{49}$$

Moreover, suppose that $\log(1/(\beta_n - \varphi_n)) \le C_2 \log n$ and $\varphi_n \log n \ge c_2$ for some constants $c_2, C_2 > 0$. Then there exist positive constants c, C depending only on c_1, C_1, c_2, and C_2 such that under H_0,

$$\mathrm{P}(T^B > c^{B,3S}(\alpha)) \le \alpha + Cn^{-c}.$$

In addition, the bound holds uniformly over all distributions $\mathcal{L}_\xi$ satisfying (42), (43), (48), and (49).

Comment 4.11 (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 $\varphi_n \log n \ge c_2$ by a weaker condition $\varphi_n \ge Cn^{-c}$ 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 $\varphi_n \log n \ge c_2$; in simulations reported in Section 6, we set $\varphi_n = \beta_n/2$.

[17] In the definition of the bootstrap test statistic $W_{\widehat{J}_B \cap \widehat{J}''_B}$, the set $\widehat{J}''_B$ is different from $\widehat{J}'_B$, which is used in the definition of the test statistic T^B. This is because our proof techniques do not allow us to show the validity of the critical values based on $W_{\widehat{J}_B \cap \widehat{J}'_B}$ since $\widehat{J}'_B$ is random. Instead, our approach consists of finding non-random set J such that with large probability, $\widehat{J}'_B \subset J \subset \widehat{J}''_B$, so that $T^B = \max_{j \in \widehat{J}'_B} \sqrt{n}\mu_j/\widehat{\sigma}_j \le \max_{j \in J} \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j$ and $W_{\widehat{J}_B \cap \widehat{J}''_B} \ge W_{\widehat{J}_B \cap J}$ and then showing validity of using $W_{\widehat{J}_B \cap J}$ to approximate the distribution of $\max_{j \in J} \sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j$.

5. Power

In this section, we discuss power properties of our tests. Consider the same general setup described in the Introduction and assume that (12) holds. Let the test statistic T be defined by (13). Pick any $\alpha \in (0, 1/2)$ and consider the test of the form

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

where $\widehat{c}(\alpha)$ is equal to $c^{SN}(\alpha)$, $c^{SN,2S}(\alpha)$, $c^{MB}(\alpha)$, $c^{MB,2S}(\alpha)$, $c^{EB}(\alpha)$, $c^{EB,2S}(\alpha)$, $c^{MB,H}(\alpha)$, or $c^{EB,H}(\alpha)$. We have the following result on the rate of uniform consistency of this test:

Theorem 5.1 (Rate of uniform consistency). *Suppose there exist constants* $0 < c_1 < 1/2$ *and* $C_1 > 0$ *such that*

$$M_{n,4}^2 \log^{1/2} p \leq C_1 n^{1/2-c_1} \text{ and } \log^{3/2} p \leq C_1 n. \tag{50}$$

In addition, suppose that $\inf_{n\geq 1}(\alpha - 2\beta_n) \geq c_1\alpha$ *whenever inequality selection is used. 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)*

$$\inf_{\mu\in\mathcal{B}_n} \mathrm{P}_\mu(T > \widehat{c}(\alpha)) \geq 1 - o(1), \tag{51}$$

where

$$\mathcal{B}_n = \left\{\mu = (\mu_1, \dots, \mu_p) : \max_{1\leq j\leq p}(\mu_j/\sigma_j) \geq \overline{r}_n = (1 + \epsilon_n)\sqrt{2(\log p_n)/n}\right\}$$

and P_μ *denotes the probability measure for the distribution* $\mathcal{L}_X$ *having mean* μ. *Moreover, the above asymptotic result (51) holds uniformly with respect to any sequence of distributions* $\mathcal{L}_X$ *satisfying (12) and (50).*

Comment 5.1 (Discussion of power properties). This theorem shows that our tests are uniformly consistent against all alternatives excluding those in a small neighborhood of alternatives that are too close to the null. As long as $p = p_n \to \infty$ as $n \to \infty$, the size of this neighborhood is shrinking at a fast rate $\sqrt{(\log p_n)/n}$. This is a fast rate because even when p is fixed, no test can be uniformly consistent against alternatives whose distance from the null converges to zero faster than $\sqrt{1/n}$. In fact, as we show in a working version of the paper,[18] when $p = p_n \to \infty$, no test can be uniformly consistent against alternatives whose distance from the null converges to zero faster than $\sqrt{(\log p_n)/n}$, and our tests are minimax optimal. Here,

[18] arXiv:1312.7614v4.

$\sqrt{\log p_n}$ is a small factor representing the cost we have to pay for dealing with a large number of inequalities.

Further, the theorem indicates that all of our tests have a fast rate of uniform consistency but it does not reveal that the bootstrap tests have better power properties than those of the SN tests. To explain, suppose for example that all inequalities are the same, that is, $X_{1j_1} = X_{1j_2}$ for all $j_1, j_2 = 1, \dots, p$ almost surely. In addition, suppose for concreteness that $\sigma = \sigma_1 = \cdots = \sigma_p = 1$. Moreover, suppose that $\mu = \mu_1 = \cdots = \mu_p$ is strictly positive but converges to zero as $n \to \infty$, that is, $\mu = \mu^n \downarrow 0$. Then the test statistic T is asymptotically equal to a $N(\sqrt{n}\mu^n, 1)$ random variable and, say, both one-step bootstrap critical values converge in probability to z_α, the $(1-\alpha)$ quantile of the $N(0,1)$ distribution. Therefore, the bootstrap tests are consistent against all alternatives such that $\sqrt{n}\mu^n \to \infty$ as $n \to \infty$. On the other hand, the one-step SN critical value is of order $\sqrt{\log p_n}$, as explained in Section 4, and the one-step SN test is only consistent against alternatives such that $\sqrt{n}\mu^n/\sqrt{\log p_n} \to \infty$. A similar discussion applies to the two-step tests. This explains the difference in power between the SN and the bootstrap tests.

Comment 5.2 (Comparison with methods for conditional moment inequalities)**.** As discussed in the Introduction, our methods can also be applied when dealing with a large number of (unconditional) moment inequalities that arise from a small number of conditional moment inequalities. Here we explain how our methods compare with those developed specifically for testing conditional moment inequalities. To fix ideas, suppose that we have one conditional moment inequality,

$$\mathrm{E}[m(Y,Z)|Z] \leq 0, \tag{52}$$

where Y and Z are random vectors and m is a known function. To transform this inequality into unconditional ones, let $w_{z,h}(Z) \geq 0$ be a positive weighting function indexed by the location point $z \in \mathcal{Z}_n$ and the bandwidth value $h \in \mathcal{H}_n$, where both $\mathcal{Z}_n$ and $\mathcal{H}_n$ are some large but finite sets. Then it follows from (52) that

$$\mathrm{E}[m(Y,Z)w_{z,h}(Z)] \leq 0, \quad \text{for all } z \in \mathcal{Z}_n \text{ and } h \in \mathcal{H}_n.$$

If (Y_i, Z_i), $i = 1, \dots, n$, is a random sample from the distribution of the pair (Y, Z), our approach would be to consider the test statistic

$$T = \max_{z \in \mathcal{Z}_n; h \in \mathcal{H}_n} \frac{n^{-1/2}\sum_{i=1}^n m(Y_i, Z_i) w_{z,h}(Z_i)}{\widehat{V}_{z,h}^{1/2}},$$

where $\widehat{V}_{z,h}$ is an estimator of $V_{z,h}$, the variance of $m(Y,Z)w_{z,h}(Z)$. This is the test statistic used in Armstrong and Chan (2016), up to a minor modification that they use infinite sets $\mathcal{Z}_n$ and $\mathcal{H}_n$. Since they couple the test statistic T with the $(1-\alpha)$ quantile of the asymptotic distribution of T when $\mathrm{E}[m(Y,Z)|Z] = 0$ almost surely, it follows that the power of their test essentially coincides with that of our one-step bootstrap tests, which can be improved by using our two-step and three-step bootstrap tests.

The approach in Chetverikov (2017), on the other hand, would be to consider the test statistic

$$T' = \max_{z\in\mathcal{Z}_n, h\in\mathcal{H}_n} \frac{n^{-1/2}\sum_{i=1}^n m(Y_i,Z_i)w_{z,h}(Z_i)}{\widehat{V}_{z,h,c}^{1/2}},$$

where $\widehat{V}_{z,h,c}$ is an estimator of $V_{z,h,c}$, the variance of $\varepsilon w_{z,h}(Z)$, where $\varepsilon = m(Y,Z) - \mathrm{E}[m(Y,Z)|Z]$. Since

$$\begin{aligned} V_{z,h} &= \mathrm{E}[m(Y,Z)^2 w_{z,h}(Z)^2] - \mathrm{E}[m(Y,Z)w_{z,h}(Z)]^2 \\ &= \mathrm{E}[(\mathrm{E}[m(Y,Z)|Z]+\varepsilon)^2 w_{z,h}(Z)^2] - \mathrm{E}[\mathrm{E}[m(Y,Z)|Z]w_{z,h}(Z)]^2 \\ &= \mathrm{Var}(\mathrm{E}[m(Y,Z)|Z]w_{z,h}(Z)) + V_{z,h,c} \geq V_{z,h,c}, \end{aligned}$$

the same alternatives will lead to larger values of T' than of T. It is therefore expected that the tests in Chetverikov (2017) would typically have better power properties than those of the tests developed in our paper.[19]

Further, it is argued in Armstrong and Chan (2016) that their test typically has better power properties than those of the test in Andrews and Shi (2013), and so, given that our methods perform at least as good as the Armstrong-Chan test, we expect that our methods also should often have better power than those in Andrews and Shi (2013), although neither approach dominates the other one. Moreover, it is important to emphasize that the Andrews-Shi test requires somewhat weaker regularity (in particular, moment) conditions than those used in our paper. Further comparisons of different methods, including those in Chernozhukov, Lee, and Rosen (2013) and in Lee, Song, and Whang (2013a,b) can be found in Chetverikov (2017).

To conclude this comparison, we emphasize that our methods are meant to complement those in the literature on testing conditional moment inequalities since our methods can be used to deal with a large number of (unconditional) moment inequalities that do not arise from the small number of conditional moment inequalities.

6. Monte Carlo Experiments

In this section, we provide results of a Monte Carlo simulation study. The simulation study consists of three parts. The first part demonstrates that the methods developed in this paper have good size control and power properties

[19] The precise comparison here is difficult. Indeed, consider for example the one-step bootstrap critical values developed here and in Chetverikov (2017). In both cases, the critical values are asymptotically equal to the $(1-\alpha)$ quantile of the maxima of $N(0,1)$ random variables, and are expected to be similar. On the other hand, the correlation structure of the $N(0,1)$ random variables in our paper and in Chetverikov (2017) are different, and so it may be possible that our tests sometimes perform better than those in Chetverikov (2017).

and also demonstrates power advantages of using bootstrap and multi-step procedures over self-normalized and one-step procedures in a broad variety of abstract settings. These abstract settings are useful because they allow us to vary the key parameters of the data-generating process in a straightforward fashion and see how the performance of our methods depend on these parameters. Importantly, this part of the simulation study shows that the size control is achieved even though we use setups with a large number of moment inequalities. The second part sheds some light on the choice of the tuning parameters for our two- and three-step methods. The third part applies our methods in an example based on the market structure model of Ciliberto and Tamer (2009).

6.1. **Size and power in abstract settings.** Throughout all the experiments in this subsection, 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.

All the experiments are based on the following data-generating process:

$$X_{ij} = \theta(1\{j \le \gamma_1 p\} + \varepsilon_{ij}) - b1\{\gamma_2 p < j \le p\} + \varepsilon_{ij}.$$

Here, θ is a scalar parameter of interest, (γ_1, γ_2, b) is a triple of additional parameters governing the data-generating process, and $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{ip})^T$, $i = 1, \dots, n$, is a sequence of i.i.d. random vectors in $\mathbb{R}^p$. We always set $\gamma_1 = 5\%$ and $\gamma_2 = 10\%$ but we vary b and the distribution of ε_i's depending on the experimental design.

We consider 8 different experimental designs. In all designs, we assume that for all $i = 1, \dots, n$, we have $\varepsilon_i = A^T \epsilon_i$, where the vector $\epsilon_i = (\epsilon_{1i}, \dots, \epsilon_{ip})^T$ consists of i.i.d. zero-mean random variables with variance one, so that the covariance matrix of ε_i's is $\Sigma = A^T A$. In Designs 1, 2, 5, and 6,

$$\Sigma_{jk} = 1\{j = k\} + \rho 1\{j \ne k\}, \quad \text{for all } j, k = 1, \dots, p.$$

In Designs 3, 4, 7, and 8,

$$\Sigma_{jk} = \rho^{|j-k|}, \quad \text{for all } j, k = 1, \dots, p.$$

We set $b = 0$ in Designs 1, 3, 5, and 7, and $b = 0.8$ in Designs 2, 4, 6, and 8. For each experimental design, we consider $\rho = 0$, 0.5, and 0.9, and we generate ϵ_{ij}'s either from Student's t distribution, which we normalize to have variance one, or from the uniform on $[-a, a]$ distribution, where we set $a = \sqrt{3}$, so that this distribution also has variance one. In the tables, where the results are presented, we write $\mathcal{L}(\epsilon) = T$ or $\mathcal{L}(\epsilon) = U$, depending on whether ϵ_{ij}'s are simulated from Student's t or from the uniform distribution.

Observe that for our data-generating process,

$$\mu_j = \mathrm{E}[X_{1j}] = \theta 1\{j \le \gamma_1 p\} - b\{\gamma_2 p < j \le p\}, \quad \text{for all } j = 1, \dots, p,$$

so that the null hypothesis (1) holds if and only if $\theta \le 0$ since we always set $b \ge 0$. We therefore consider testing (1) against (2) for $\theta = 0$ (Designs

1-4; the null holds) and $\theta = 0.07$ (Designs 5-8; the null does not hold; the value 0.07 is chosen to make sure that most probabilities are bounded away from 0 and 1). Note also that when we set $\theta = 0.07$, only $\gamma_1 = 5\%$ of the inequalities violate the null hypothesis. Moreover, when we set $b = 0.8$, $1 - \gamma_2 = 90\%$ of inequalities satisfy the null and are not binding.

We consider self-normalized (SN), multiplier bootstrap (MB), and empirical bootstrap (EB) critical values. For all three methods, we consider their one- and two-step versions. For the MB and EB methods, we also consider their three-step versions. In all experiments, we set the nominal level of the test $\alpha = 5\%$ and for the tests with the inequality selection, we set $\beta = 0.1\%$. For the three-step methods, we set $\varphi = \beta/2$. We present results based on 1000 simulations for each design, and we use $B = 1000$ bootstrap samples for each bootstrap procedure.

In addition, to see if the methods developed specifically for testing conditional moment inequalities can be used in our setting (with "unstructured" inequalities), we also consider the Andrews-Shi test (note that their approach consists of first transforming the conditional moment inequalities into many unconditional ones and then testing the unconditional moment inequalities but implementing the second step does not require knowing the original structure of the conditional moment inequalities, which makes it possible to apply their test in our setting).[20] To implement their test, we use the test statistic T' in (15), which corresponds to their CvM statistic, and obtain the critical value via a bootstrap procedure as described in Section 9 of Andrews and Shi (2013), which corresponds to their GMS critical value. We follow all their recommendations regarding the choice of the tuning parameters.

Results on the probabilities of rejecting the null in all the experiments are presented in Tables 1-4 in the online supplement. In these tables, we use B_j for $B \in \{SN, MB, EB\}$ and $j \in \{1, 2, 3\}$ to denote j-step B test. We also use AS to denote the Andrews-Shi test.

The first observation to be taken from these tables is that the MB and EB methods give similar results. The second observation is that although the Andrews-Shi test performs well in many settings, it does not control size in some settings; for example, when $p = 1000$ and $\rho = 0$, the AS test rejects the null with probability around 15% in Design 1 (Table 1), even though the null holds and the nominal level of the test is 5%. Therefore, in what follows, we only discuss and compare our SN and bootstrap (MB and EB) methods.

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

[20] The tests of Armstrong and Chan (2016) and of Chetverikov (2017) can not be implemented in our setting because they require knowledge of the original structure of the conditional moment inequalities. In particular, the critical value for the Armstrong-Chan test depends on the volume of the support of the conditioning variable and the test statistic for Chetverikov's test depends on certain conditional heteroscedasticity functions.

occurs in Design 3 with autocorrelated data, uniform ϵ_{ij}'s, $p = 500$, and $\rho = 0$, where the one-step EB test rejects the null with probability 7.7% against the nominal level $\alpha = 5\%$ (Table 2). As expected, the self-normalized tests 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 reject the null with probability between 4.8% and 5.2%, which is very close to the nominal level $\alpha = 5\%$, but both the SN tests never reject the null. 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 $\theta = 0.07$ and H_0 does not hold, and demonstrate power properties of our tests. Note that we have for all $j = 1, \dots, p$ that $\mathrm{Var}(X_{1j}) = (1 + \theta)^2 = 1.07^2 = 1.1449$. Hence, if we had only one inequality to test ($p = 1$), non-trivial testing would only be possibly for μ_1 at least of order $(1.1449/n)^{1/2} = 1.07/20 = 0.0535$. Instead, we have many inequalities (p is large) but we set $\mu_j = 0.07$ for the inequalities that violate the null, which is of the same order as 0.0535. Note also that in our setting, only $\gamma_1 = 5\%$ of all inequalities violate the null. Therefore, since Tables 3 and 4 show that our methods yield non-trivial rejection probabilities in most cases and sometimes yield the rejection probability close to one, we conclude that our methods have good power properties. The one-step and two-step SN tests have rejection probabilities close to those for the corresponding bootstrap tests when $\rho = 0$ or even when $\rho = 0.5$ for Designs 7 and 8 with autocorrelated data. Further, the one-step and two-step bootstrap tests substantially improve upon the corresponding SN tests 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 around 20% and the corresponding bootstrap tests reject H_0 with probability around 40%. Finally, selection procedures yield important power improvements. For example, for Design 8 with autocorrelated data, ϵ_{ij} having Student's t-distribution, $p = 1000$ and $\rho = 0.5$, the one-step MB method reject the null with probability around 40% but the two-step method reject with probability around 90%. Similarly, In Design 7 with autocorrelated data, ϵ_{ij} having the uniform distribution, $p = 200$ and $\rho = 0$, the two-step EB method rejects the null with probability around 50% and the three-step EB method rejects with probability around 80%.

6.2. **Selecting tuning parameters.**

In this subsection, we carry out a small simulation study to develop a rule of thumb for selecting the tuning parameters for our methods. Since the bootstrap methods are more powerful than the SN methods, we do not consider the SN methods here. Also, since the MB and EB methods give similar results, we focus on the MB methods only. Thus, in this subsection, we only discuss the two-step and three-step MB methods but note that the same discussion applies to the corresponding EB methods.

We consider the same data-generating process as that in Design 5 in the previous subsection with $\rho = 0$, ϵ_{ij}'s having uniform distribution, and $\theta = 0.07$. Instead of setting $b = 0$, however, we vary b from 0.05 to 0.8 to see how it affects the choice of the tuning parameters. We consider both the two-step and the three-step MB methods with $\alpha = 5\%$ and β varying from 0.1% to 1.0%. For the three-step MB method, we set $\varphi = \beta/2$. Depending the simulation, we set $p = 200$ or 1000. As in the previous subsection, we present results based on 1000 simulations for each setting, and we use $B = 1000$ bootstrap samples for each bootstrap procedure. In unreported simulations, we also tried to vary ρ and to use Student's distribution for ϵ_{ij}'s and found results similar to those reported below. Results for the two-step and three-step MB methods are presented in Tables 5 and 6, respectively, in the online supplement.

Before looking at the simulation results, we provide some intuition regarding the choice of the tuning parameters. First, we discuss the two-step MB method, which requires selecting the tuning parameter β. Observe that increasing β has two effects on the power of the method. One effect is that holding $\widehat{J}_{MB}$ fixed, increasing β leads to higher values of $c^{MB,2S}(\alpha)$ since $c^{MB,2S}(\alpha)$ is defined as the $(1 - \alpha + 2\beta)$-quantile of the conditional distribution of $W_{\widehat{J}_{MB}}$ given X_1^n; see (38). The other effect is that increasing β shrinks the set $\widehat{J}_{MB}$, which is defined as the set of all j's such that $\sqrt{n}\widehat{\mu}_j/\widehat{\sigma}_j > -2c^{MB}(\beta)$. This in turn leads to smaller values of $c^{MB,2S}(\alpha)$. Since the test statistic T does not depend on β, the first effect decreases the power of the method and the second one increases it. Selecting β therefore requires balancing these two effects.

Further, observe that the second effect is negligible when all inequalities satisfying the null are binding or nearly binding since these inequalities will be in the set $\widehat{J}_{MB}$ even for large values of β. Similarly, the second effect is negligible when all inequalities satisfying the null are far away from being binding since these inequalities will be out of the set $\widehat{J}_{MB}$ even for small values of β. Thus, the second effect is non-negligible, so that it might be useful to use large values of β, only when there are inequalities under the null that are not too close and not too far away from being binding.

Our simulation results support the discussion above. Indeed, as follows from Table 5, for $p = 200$, the power of the two-step MB method is a decreasing function of β when $b < 0.40$ and when $b > 0.55$. Therefore, the second effect is strong enough to create a non-monotonicity in the power function only in a small range of the values of b. Even in these cases, however, the second effect is not strong enough, so that setting $\beta = 0.1\%$

yields almost the same power as the power we would obtain by selecting β optimally. Similar discussion also applies when $p = 1000$. Hence, the simulation results in Table 5 suggest that setting $\beta = 0.1\%$ is a good rule of thumb.[21]

Next, consider the three-step MB method. The problem of selecting the tuning parameters is now much more complicated because we now have to choose two parameters, β and φ. Regarding the choice of φ, for given value of β, selecting φ exhibits a trade off between good power and size control: choosing larger values of φ improves the size control but undermines the power of the test. Since there are no universally accepted rules in the literature on striking the balance between power and size control, and since our results (Theorem 4.6) require that φ is not too close to zero and not too close to β, we simply set $\varphi = \beta/2$. Regarding the choice of β, although the situation is now more difficult relative to what we had with the two-step method because now both the test statistic and the critical value depends on β, the overall trade off is similar to what we had before. In particular, the simulation results in Table 6 reveal that the power of the three-step MB method is always a decreasing function of β. We therefore, again, conclude that setting $\beta = 0.1\%$ is a good rule of thumb.

6.3. **An application to market structure model.** In this subsection, we show how our methods apply in an economic model setting. Specifically, we consider the market structure model from Section 2. For a given market, three firms ($m = 3$) are simultaneously deciding whether to enter the market or not. For $j = 1, \dots, 3$, let $D_j = 1$ if the firm j enters the market and $D_j = 0$ otherwise. If the firm j enters the market, its profit is given by

$$\pi_j = \sum_{l \neq j} \delta_{lj} D_l + \varepsilon + \zeta_j,$$

where ε is the market size shock that is common to all three firms, and ζ_j is an idiosyncratic shock representing specific conditions of the firm j in the market. If the firm j does not enter the market, $\pi_j = 0$. The objective of each firm is to maximize its profit given the decisions of other firms.

We assume that ε, ζ_1, ζ_2, and ζ_3 are i.i.d. standard normal random variables. The parameter δ_{lj} represents the effect of the presence of the firm l in the market on the firm j. To simplify the setting, we assume that $\delta_{lj} = \delta_{jl}$ for all $j, l = 1, \dots, 3$ with $j \neq l$, so that the firms have symmetric effects on each other. With this assumption, we use the following reparameterization of the model:

$$\theta_1 = \delta_{12}, \quad \theta_2 = \delta_{13}, \quad \theta_3 = \delta_{23}.$$

[21] Note also that it is almost never useful to set $\beta < 0.1\%$ since in this case, holding $\widehat{J}_{MB}$ fixed, we would obtain essentially the same critical value $c^{MB,2S}(\alpha)$ as the one given by $\beta = 0.1\%$, but the substantial cost of setting $\beta < 0.1\%$ is that it can significantly increase the set $\widehat{J}_{MB}$, relative to the set we obtain by setting $\beta = 0.1\%$.

The random variables ε, ζ_1, ζ_2, and ζ_3 are observed by the firms when they make their decisions but are not observed by the researcher. For simplicity, we also assume away any variation X that is observed by the researcher.

We assume that the parameters θ_1, θ_2, and θ_3 are all negative, so that the game always has a Nash equilibrium in pure strategies, and we focus on such equilibria. When there is only one equilibrium, we assume that the outcome of the game $D = (D_1, D_2, D_3)$ is determined by this equilibrium. When there are several equilibria, we assume that the outcome is determined by a randomly selected equilibrium, where all equilibria have the same probability of being chosen.

We consider inference on the parameters θ_1, θ_2, and θ_3 using the data on market outcomes for n i.i.d. markets. If the researcher knew that the outcome of the game were determined by a randomly selected equilibrium whenever there are several equilibria, the model would be point identified, and there would be only one value of the parameters consistent with the distribution of the outcomes. However, since the researcher typically has no reasons to believe that a particular equilibrium selection mechanism is used, we consider inference approaches from the literature on partial identification, which are agnostic about the equilibrium selection mechanism.

Specifically, we consider two types of bounds: the Ciliberto and Tamer (2009) bounds and the Galichon and Henry (2011) bounds. The Ciliberto-Tamer (CT) bounds, which are described in Section 2, give $2 \cdot 2^m = 2 \cdot 2^3 = 16$ inequalities:

$$P_1(d, \theta) \leq \mathrm{E}[1\{D = d\}] \leq P_2(d, \theta), \text{ for all } d \in \mathcal{D}, \tag{53}$$

where $\mathcal{D} = \{0,1\}^m = \{0,1\}^3$ is the set of all possible outcomes, $P_1(d, \theta)$ is the probability that the outcome d is the unique equilibrium of the game, and $P_2(d, \theta)$ is the probability that the outcome d is an equilibrium of the game. Since the probabilities $P_1(d, \theta)$ and $P_2(d, \theta)$ are hard to calculate exactly, we approximate them numerically using 100000 simulations of the game.

To describe the Galichon-Henry (GH) bounds, for each set of outcomes $A \subset \mathcal{D}$, let $\mathcal{L}(A, \theta)$ be the probability of observing an outcome in A under the assumption that whenever the game has several equilibria, some of which are in A and others are not, an equilibrium from A is selected. Then the GH bounds give inequalities

$$\mathrm{E}[1\{D \in A\}] \leq \mathcal{L}(A, \theta), \text{ for all } A \subset \mathcal{D}. \tag{54}$$

Thus, for each set A, we get one inequality, and so in total we obtain $2^{|\mathcal{D}|} = 2^{2^3} = 2^8 = 256$ inequalities. Note, however, that when $A = \emptyset$, the empty set, or $A = \mathcal{D}$, we obtain inequalities that always hold, and so we can disregard them. Thus, we have $256 - 2 = 254$ inequalities.

The major advantage of the GH bounds is that they are tight and yield the sharp identified set for $\theta = (\theta_1, \theta_2, \theta_3)$, that is, it is never possible, without further assumptions, to find a value of θ that would satisfy the inequalities

(54) but would be inconsistent with the distribution of the outcomes of the game. The CT bounds do not necessarily have this property, and it may be possible to find a value of θ that would satisfy (53) but would not satisfy (54). On the other hand, even though the GH bounds are useful for *the identification analysis*, since they produce a lot of inequalities even in simple models (254 in our case, which is a large number, and our game has only three firms), it was previously not possible to use them for *inference* on θ. This is, however, possible using our methods. We are therefore interested to see, via simulations, how the GH bounds work for inference and also to compare the inference based on the GH bounds with that based on the CT bounds.

For our simulations, we consider samples of size $n = 1000$, 2000, and 5000, which are comparable with the sample size in Ciliberto and Tamer (2009), $n = 2742$. We always set $\theta_1 = -0.6$ and $\theta_2 = \theta_3 = -1.3$, and we consider testing the null hypothesis $H_0 : \theta = \theta_0$ for different values of $\Delta\theta = \theta_0 - \theta$. To investigate size control of our methods, we use $\Delta\theta = (0, 0, 0)$, and to investigate their power, we use $\Delta\theta = (0.25, 0, 0)$, $(-0.25, 0, 0)$, $(0, 0.25, 0)$, $(0, -0.25, 0)$, $(0, 0, 0.25)$, and $(0, 0, -0.25)$. We consider the one-step and two-step versions of the SN, MB, and EB methods. In addition, we consider the three-step versions of the MB and EB methods. Note, however, that the market structure model studied here violates the conditions required for our three-step methods. In particular, we require in Section 4.4 that the gradients (with respect to the parameters) of the moment functions have non-vanishing variance, $\sigma^V_{jl} > 0$, but the corresponding gradients here are non-stochastic and so have variance zero. Therefore, as a way to drop weakly informative inequalities in the three-step methods, we drop all the inequalities that have $|\mu^V_{jl}| \leq 1/\sqrt{n}$ for all $l = 1, 2, 3$ in the notation of Section 4.4. We tried replacing $1/\sqrt{n}$ by $0.5/\sqrt{n}$ and $2/\sqrt{n}$ but obtained similar results. For all methods, we set $\alpha = 5\%$ and whenever needed, $\beta = 0.1\%$. For all bootstrap methods, we use 500 bootstrap samples, and for each simulation design, we repeat the experiment 1000 times to obtain rejection probabilities. The results of our simulation study are presented in Table 7 in the online supplement.

Table 7 shows that all of our methods have good size control. In particular, when $\Delta\theta = (0, 0, 0)$, the rejection probabilities do not exceed 3.8%. Also, the GH bounds give somewhat more conservative results in comparison with the CT bounds. Regarding the power, it is important to note that since the market structure model is partially identified, our methods have relatively low power against some alternatives (for example, $\Delta\theta = (0, 0, -0.25)$) even when $n = 5000$ (no methods may have power against θ_0 in the sharp identified set). The MB and EB methods give similar results, and the bootstrap methods are more powerful than the SN methods, especially in the case of the GH bounds; for example, when $\Delta\theta = (0, 0.25, 0)$ and $n = 5000$, the two-step MB method based on the GH bounds rejects the null with probability

53% whereas the corresponding two-step SN method rejects the null with probability 36%. Three-step methods give results similar to those for the two-step methods.

Further, it is intuitively clear that in comparison with the CT bounds, the GH bounds may be much more powerful against those θ_0 that satisfy or nearly satisfy (53) but do not satisfy (54). This can be seen for $\Delta\theta = (-0.25, 0, 0)$ and $n = 5000$, where the two-step MB method based on the GH bounds rejects the null with probability 99% but the same method based on the CT bounds rejects the null with probability only 70% (in fact, as was reported in the previous version of the paper, when we set $\Delta\theta = (-0.2, 0, 0)$ and $n = 5000$, the two-step MB method rejects the null with probability 87% when the GH bounds are used and only 18% when the CT bounds are used). This is an important advantage of the GH bounds. On the other hand, whenever θ_0 does not satisfy (53), the methods based on the CT bounds may be more powerful because they use a smaller set of inequalities, and the critical values for our methods are increasing with the number of moment inequalities used. However, the simulation results reveal that the methods based on the GH bounds, even though sometimes less powerful, are always comparable with those based on the CT bounds. When the two-step MB method is used, perhaps the largest difference in power occurs for $\Delta\theta = (0, -0.25, 0)$ and $n = 5000$, where the CT and GH bounds yield the rejection probabilities 48% and 34%, respectively.

References

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

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. (2015). Asymptotically exact inference in conditional moment inequality models. *Journal of Econometrics* **186** 51-65.

Armstrong, T.B. and Chan, H.P. (2016). Multiscale adaptive inference on conditional moment inequalities. *Journal of Econometrics* **194** 24-43..

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. (2013). 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. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields.* **162** 47-70.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309-2352.

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

Chetverikov, D. (2017). Adaptive test of conditional moment inequalities. *Econometric Theory* 1-42.

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

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.

Hotz, V. and Miller, R. (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies* **60** 497-529.

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.

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.

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.

Pearl, J. (2009). *Causality: models, reasoning, and inference.* Cambridge University Press.

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. (2014). A practical two-step method for testing moment inequalities with an application to inference in partially identified models. *Econometrica* **82** 1979-2002.

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.

Shah, R. and Peters, J. (2018). The hardness of conditional independence testing and the generalised covariance measure. arXiv:1804.07204.

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

White, H. (2000). A reality check for data snooping. *Econometrica* **5** 1007-1126.

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