Readings on Moment Inequality Methods

Manu Navjeevan

June 8, 2020

Contents

1 A Practical Method for Testing Many Moment Inequalities; *Yuehao Bai, Andres Santos, Azeem M. Shaikh*

1.1 Introduction

Setup: $\{X_i\}_{i=1}^n$ i.i.d with distribution $P \in P_n$ on R^n. Consider the problem of testing

$$H_0 : P \in \mathbf{P}_{0,n} \quad \text{versus} \quad H_1 : P \in \mathbf{P}_{1,n} \tag{1}$$

where

$$\mathbf{P}_{0,n} \equiv \{P \in \mathbf{P}_n : E_P[X_i] \leq 0\} \tag{2}$$

and $\mathbf{P}_{1,n} = \mathbf{P}_n/\mathbf{P}_{0,n}$. The inequality in 2 is interpreted component wise and $\mathbf{P}_n$ is a large class of possible distributions for the observed data. Indexing both the number of moments p_n and the class of possible distributions by the sample size allows for the number of moments to grow (rapidly) with the sample size n. Goal is to construct test that are uniformply consistent in level; i.e

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_{0,n}} E_P[\phi_n] \leq \alpha \tag{3}$$

A test can be viewed as a function of the data $\phi_n = \phi_n : \mathcal{X}^n \to \{0, 1\}$ where $\mathcal{X}^n$ is generally some subset of $\mathbb{R}^n$ where the data takes its values.

There are a large class of problems in economics in which the number of moments is large. For example, in the entry models as in Cilberto and Tamer (2009) the number of moment inequalities to check is $p_n = o(2^{m+1})$ where m is the number of firms. Apart from Chernozukhov et. al (2019), this has typically been done by limiting $\mathbf{P}_n$ so that the number of moments p_n are small. Canay and Shaikh (2017) provide a detailed review of these tests. This paper focuses on the two step testing procedure of Romano et. al (2014). Test is shown to satisy (3) under assumptions on $\mathbf{P}_n$ that restrict p_n to not depend on n. However, the test is "practical" in that it is computationally feasible even if the number of moments is large. **Paper shows that the test of Romano et. al (2014) continues to satisy (3) for a large class of distributions that permits the number of moments p_n to grow exponentially with the sample size n.**

Theoretical analysis relies on Chernozhukov et. al (2013, 2017) on the high dimensional CLT. This is seminal work. Allen (2018) argues that the test proposed Romano et al. (2014) is more powerful in finite samples then the test proposed by Chernozhukov et al. (2019).

1.2 Main Result

Begin this section by describing the testing procedure in Romano et al. (2014). To do so, best to introduce some further notation. For $1 \leq j \leq p_n$ let $X_{i,j}$ denote the jth component of X_i and set

$$\bar{X}_{j,n} \equiv \frac{1}{n}\sum_{i=1} nX_{i,j} \tag{4}$$

$$S_{j,n}^2 \equiv \frac{1}{n}\sum_{i=1}^{n}(X_{i,j} - \bar{X}_{j,n})^2 \tag{5}$$

Can also use the notation $\mu_j(p) \equiv E_P[X_{i,j}]$ and $\sigma_j^2(P) \equiv \mathrm{Var}_P[X_i, j]$ so that (4) and (5) can be expressed as $\mu_j(\hat{P}_n)$ and $\sigma_j^2(\hat{P}_n)$, respectively, where $\hat{P}_n$ is the empirical distribution of $\{X_i\}_{i=1}^n$. Focus on a test that rejects for large values of

$$T_n \equiv \max\left\{\max_{1\leq j\leq p_n} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}}, 0\right\}$$

In defining critical value, useful to introduce an i.i.d sequence of random variables with distribution $\hat{P}_n$ conditional on $\{X_i\}_{i=1}^n$, which we will denote $X_i^*, i = 1, \ldots, n$. Further define $\bar{X}_{j,n}^*$ and $(S_{j,n}^*)^2$ analagously to before, but substituting in X_i^*. Critical value for T_n is given by

$$\hat{c}_n^{(2)}(1-\alpha+\beta) \equiv \inf \mathcal{S}_n(1-\alpha+\beta) \tag{6}$$

where

$$\mathcal{S}_n(a) \equiv \left\{ c \in \mathbb{R} : \mathbb{P}\left[\max\left\{ \max_j \frac{\sqrt{n}(\bar{X}_{j,n}^* - \bar{X}_{j,n} + \hat{\mu}_{j,n})}{S_{j,n}^*}, 0 \right\} \leq c \;\middle|\; \{X_i\}_{i=1}^n \right] \geq a \right\}$$

Here $\alpha \in (0, 0.5)$ is the nominal level of the test and $\beta \in (0, \alpha)$ and

$$\hat{\mu}_{j,n} \equiv \min\left\{ \bar{X}_{j,n} + \frac{S_{j,n}}{\sqrt{n}} \hat{c}_n^{(1)}(1-\beta), 0 \right\} \tag{7}$$

with

$$\hat{c}_n^{(1)} \equiv \inf\left\{ c \in \mathbb{R} : \mathbb{P}\left[\max_{1 \leq j \leq p_n} \frac{\sqrt{n}(\bar{X}_{j,n} - \bar{X}_{j,n}^*)}{S_{j,n}^*} \leq c \;\middle|\; \{X_i\}_{i=1}^n \right] \geq 1-\beta \right\}$$

The test is then

$$\phi_n^{\text{RSW}} \equiv \mathbb{1}\left\{ T_n \geq \hat{c}_n^{(2)}(1-\alpha+\beta) \right\} \tag{8}$$

Motivating this choice of critical value it is useful to note that the test statistic T_n satisfies

$$T_n = \max\left\{ \max_j \left(\frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P))}{S_{j,n}} + \frac{\sqrt{n}\mu_j(P)}{S_{j,n}} \right), 0 \right\} \tag{9}$$

Decomposition highlights that the main impedment in approximating the distribution of T_n is the presence of nuisance parameters $\sqrt{n}\mu_j(P)$ for $1 \leq j \leq p_n$.[1] Though these nuisance parameters cannot be consistently estimated, Romano et al (2014) observe that it may still be possible to construct a suitably valid confidence region for them.

Lemma in Appendix employs Romano insight and high dimensional CLT of Chernozhukov et al. (2017) to show that, under conditions that permit p_n to grow rapidly with the sample size n, $\sqrt{n}\mu_j(P) \leq \sqrt{n}\hat{\mu}_{j,n}$ for all $j \leq p_n$ with pr. approximately no less than $1-\beta$ whenever the null hypothesis in (1) is true. Since T_n is monotonically increasing in the nuisance parameters $\sqrt{n}\mu_j(P)$ for all $1 \leq j \leq p_n$ it follows that, viewed as a funtion of these nuisance parameters, any quantile of T_n is maximized over said confidence region by setting $\sqrt{n}\mu_j(P) = \sqrt{n}\hat{\mu}_{j,n}$ for all j. Then, the critical value $\hat{c}_n^{(2)}(1-\alpha+\beta)$ is a bootstrap estimate of the $1-\alpha+\beta$ quantile of T_n under the "least favorable" nuisance parameter value $\sqrt{n}\mu_j(P) = \sqrt{n}\hat{\mu}_{j,n}$ for all j. The $1-\alpha-\beta$ quantile is employed instead of β to account for that, with pr. appx no greater than β, $\sqrt{n}\mu_j(P) > \sqrt{n}\hat{\mu}_{j,n}$. Analysis of test (8) hinges on following assumption:

Assumption 1. Assume (i) $\{X_i\}_{i=1}^n$ is an i.i.d sample with $X_i \in \mathbb{R}^{p_n}$ and $X_i \sim P \in \mathbf{P}_n$; (ii) $\sigma_j(P) > 0$ for all $1 \leq j \leq p_n$ and $P \in \mathbb{P}_n$; (iii) For $k = 1, 2$, there is a $M_{k,n} < \infty$ such that $E_P[|X_{i,j} - \mu_j(P)|^{2+k}] \leq \sigma_j^{2+k}(P) M_{k,n}^k$ for all $1 \leq j \leq p_n$ and $P \in \mathbf{P}_n$; (iv) There exists a $B_n < \infty$ such that $E_P\left[\max_{1 \leq j \leq p_n} \left|X_{i,j} - \mu_j(P)\right|^4\right] \leq B_n^4$ for all $P \in \mathbf{P}_n$; (v) $(M_{1,n}^2 \vee M_{2,n}^2 \vee B_n^2)\log^{3.5}(p_n n) = o(n^{(1-\delta)/2})$ for some $\delta \in (0,1)$

1(i) formalizes that $\{X_i\}_{i=1}^n$ be an i.i.d sample, while Assumption 1(ii) requires the variance of $X_{i,j}$ to be positive for all $P \in \mathbb{P}_n$ and $1 \leq j \leq p_n$.1(iii) imposes a uniform in P and j bound on the standardized moments of $X_{i,j}$. Condition is a strengthening of the uniform integrability requirements of Romano et al (2014) required so study a setting in which p_n diverges to infinity. Part (iv) bounds the 4th moments of the maximim of $X_{i,j}$. Finally, (v) states the main conditiion governing how fast p_n can grow with n. Under suitable moment restrictions on $X_{i,j}$, p_n may grow exponentially with n. Now ready for main result

[1] I'm not entirely sure why they cannot be consistently estimated. I think this is because we are only partially identified.

Theorem 1. *If Assumption 1 holds, $\alpha \in (0, \frac{1}{2})$ and $0 < \beta < \alpha$, then ϕ_n^{RSW} as defined in (8) satisfies uniform consistency in level as defined in (3)*

The rest of this paper goes through some simulations. It is also just a working paper at the moment. Probably it is best to go through the main proof; but I will print it out and make some notes on this.

2 Inference on Causal and Structural Parameters Using Many Moment Inequalities; *Victor Chernozhukov, Denis Chetverikov, and Kengo Kato (ReStud, 2019)*

2.1 Introduction

In recent years, moment inequalities framework has developed into a powerful tool for inference on causal and structural parameters in partially identified models. Many papers study models with a finite and fixed number of conditional and unconditional moment inequalities. IN practice the number of moment inequalities implied by the model is often large.

Examples of testing (very) many moment inequalities

- Consumer is selecting a bundle of products for purchase and moment inequalities come from revealed preference argument (Pakes, 2010)
- Market structure model of Ciliberto and Tamer (2009), number of moment inequalities equals the number of possible combinations of firms that could potentially enter the market (grows exponentially in the number of firms)
- Dynamic model of imperfect competition of Bajari, Benkard, Levin (2007)m where deviations from optimal policy serve to define many moment inequalities
- Beresteanu, Molchanov, Molinari (2011), Galichon and Henry (2011)[2], Chesher, Rosen, Smolinksi (2013), and Chester and Rosen (2013)

Many examples have inportant in that the many inequalities under consideration are "unstructured", they cannot be viewed as unconditional moment inequalities generated froma small number of conditional inequalities with a low-dimensional conditioning variable. So existing inference methods for conditional moment inequalities, though fruitful in many cases

Formally describing the problem, let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d random vectors in $\mathbb{R}^p$, where $X_i = (X_{i1}, \ldots, X_{ip})^T$, with a common distribution denoted by $\mathcal{L}_X$. For $j \leq p$, we write $\mu_j := \mathbb{E}[X_{1j}]$. Interested in testing the null hypothesis

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

Against the alternative

$$H_1 : \mu_j > 0 \;\; \text{for some } j = 1, \ldots, p \tag{2}$$

Refer to (1) as the moment inequalities and say the jth moment is satisfied (viloated) if $\mu_j \leq 0$ ($\mu_j > 0$). Paper will allow number of moment inequalities $p \gg n$. Consider a test statistic given by the maximum over p Studentized (t-type) inequality specific statistic. Consider critical values based upon (i) the union bound combined with a moderate deviation inequality for self-normalized sums and (ii) bootstrap methods. Among bootstrap methods, consider multiplier and emppirical bootstrap methods. These are simulation based and computationally more difficult, but take into account correlation structure and yeild lower critical values. SN method is particularly useful for grid search when the researcher is interested in constricting a confidence interval for identified set.

Also consider two-step methods incorporating inequality selection procedutes. Two-step methods get rid of most uninformative inequalities, that is inequalities with $\mu_j < 0$ if μ_j is not too close to 0. Also develop novel three-step methods by incorporating double inequality selection procedures. These are suitable in parametric modesl defined via momoent inequalities and allow to drop weakly informative inequalities in

[2] This seems like a good place to start reading

addition to uninformative inequalities.[3]. Results can be used for construction of confidence regions for identifiable parameters in partially identified models defined by moment inequalities. Show that results are asymptotically honest (don't quite know what this means).

Literature 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).

In this paper we implicitly assume that $X_1, \dots, X_n$ and p are indexed by n. Mainly interested in the case that $p = p_n \to \infty$ as $n \to \infty$

2.2 Motivating Examples

Section provides 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 inequalities. Results cover conditioning as well.

2.2.1 Market Structure Model

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 possible values of D. We have that $|\mathcal{D}| = 2^m$.

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

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

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

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

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 $\epsilon \in D_1(d, X, \theta)$ and d is *an* equilibrium whenever $\epsilon \in R_2(d, X, \theta)$. In the second case, the probability that the researcher sees d as an equilibrium depends on the equilibrium selection mechanism. Without further information, anything can be in $[0, 1]$. Therefore we have the following bounds

$$\begin{aligned}\mathbb{E}[\mathbb{1}\{\epsilon \in R_1(d, X, \theta\}|X] &\leq \mathbb{E}[\mathbb{1}\{D = d\}|X] \\ &\leq \mathbb{E}[\mathbb{1}\{\epsilon \in R_1(d, X, \theta) \cup R_2(d, X, \theta)\}|X]\end{aligned}$$

Further assuming that the conditional distribution of ϵ given X is known (or known up to a parameter that is part of θ), both the LHS and RHS of these inequalites can be calculated. Denote them $P_1(d, X, \theta)$ and $P_2(d, X, \theta)$, respectively to obtain

$$P_1(d, X, \theta) \leq \mathbb{E}[\mathbb{1}\{D = d\}|X] \leq P_2(d, X, \theta) \tag{3}$$

for all $d \in \mathbb{D}$. These can be used for inference on the parameter θ. Note that the number of inequalities in (3) is $2|\mathcal{D}| = 2^{m+1}$. This is a large number, even if m is moderately large. Moreover, these inequalities are conditional on X. So, they can be transformed into a large and increasing number of unconditional moment inequalities as described above. Also, if the firms have more than two decisions, the number of inequalities will be even larger.

Some other examples are given, but I won't cover them in notes.

[3] Can be extended to nonparametric models as well

2.3 Test Statistic

Begin preparing some notation. Assume that

$$\mathbb{E}[X_{1,j}^2] < \infty, \sigma_j^2 := \mathrm{Var}(X_{1,j}) > 0, j = 1, \dots, p \tag{4}$$

For $j = 1, \dots, p$ let $\hat{\mu}_j$ and $\hat{\sigma}_j$ be the sample mean and variance of $\{X_{i,j}\}_{i=1}^n$. Many different possible test statistics. Somewhat natural to consider statistics that take large values thehn some of $\hat{\mu}_j$'s are large. In this paper focus on statistic that takes large values when at least one of $\hat{\mu}_j$ are large.

In specific, focus on the following test statistic:

$$T = \max_{1 \le j \le p} \frac{\sqrt{n}\, \hat{\mu}_j}{\hat{\sigma}_j} \tag{5}$$

Large values of T indicate a likely violation of H_0, so it is natural to consider tests of the form

$$T > c \implies \text{ reject } H_0$$

where c is appropriately chosen so that the test approximately has size $\alpha \in (0, 1)$. Consider various ways for calculating critical values and prove their validity.

2.4 Critical Values

Now move to define critical values for T such that under H_0, the probability of rejecting H_0 does not exceed size α asymptotically. Methods are ordered bu increasing computational complexity, increasing strength of required conditions, and also increasing power. Basic idea for the construction of critical values for T lies in the fact, that, under H_0:

$$T \le \max_{1 \le j \le p} \frac{\sqrt{n}(\hat{\mu}_j - \mu_j)}{\hat{\sigma}_j}$$

Consider two approaches to constructing such critical values: self-normalized and bootstrap methods. Also consider two- and three-step variants of the methods by incorporating inequality selection.

Following notation used:

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

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

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

The dependence on n comes via the dependence of $p = p_n$ on n implicitly. By Jensen's inequality, $B_n \ge M_{n,4} \ge M_{n,3} \ge 1$. In addition, if all Z_{ij}'s are bounded a.s by a constant C, we have that $C \ge B_n$. These are useful to get a sense of various conditions on $M_{n,3}, M_{n,4}$ and B_n imposed in the theorems below.

2.4.1 Self Normalized Critical Values

One-step method: Self-normalized method considered is based on the union bound combined with moderate deviation inequality for self-normalized sums. Under H_0

$$\mathbb{P}(T > c) \le \sum_{j=1}^{p} \mathbb{P}(\sqrt{n}(\hat{\mu}_j - \mu_j)/\, \hat{\sigma}_j > c) \tag{6}$$

This bound seems crude when p is large. However, will exploit the self normalizing $\sqrt{n}(\hat{\mu}_j - \mu_j)/\, \hat{\sigma}_j$ to show that RHS of above is bounded, even if c is growing logarithmically fast with p. Using such a c will yield a test with better power properties.

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

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

Simple algebra yields, we see that

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

where the right-hand side is increasing in U_j as long as $U_j \geq 0$. So under H_0,

$$\mathbb{P}(T > c) \leq \sum_{j=1}^{p} \mathbb{P}\left(U_j > c/\sqrt{1 + c^2/n}\right), \ c \geq 0 \tag{7}$$

Moderate deviation inequality for self-normalized sums of Jing, Shao, and Wang (2003) implies that for moderatley large $c \geq 0$,

$$\mathbb{P}\left(U_j > c/\sqrt{1 + c^2/n}\right) \approx \mathbb{P}\left(Z > x/\sqrt{1 + c^2/n}\right)$$

where $Z \sim N(0, 1)$. The above approximation holds even if Z_{ij} only have $2 + \delta$ finite moments for some $\delta > 0$. Therefore, take the critical value as

$$c^{SN}(\alpha) = \frac{\Phi^{-1}(1 - \alpha/p)}{\sqrt{1 - \Phi^{-1}(1 - \alpha/p)^2/n}} \tag{8}$$

where $\Phi(\cdot)$ is the normal cdf. We call $c^{SN}(\alpha)$ the one-step SN critical value with size α as it's derivation depends on the moderate deviation inequality for self-normalized sums. Note that

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

so $c^{SN}(\alpha)$ depends on p only through $\log(p)$. Following theorem provides a non asymptotic bound on the probability that the rest statsistic T exceeds the SN critical value $c^{SN}(\alpha)$ under H_0 and shows that the bound converged to α under mild regularity conditions, validating the SN method.

Theorem 2. *(Validity of one-step SN method). Suppose that $M_{n,3}\Phi^{-1}(1 - \alpha/p) \leq n^{1/6}$. Then under H_0,*

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

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

$$M_{n,3}^3 \log^{3/2}(p/\alpha) \leq C_1 n^{1/2 - c_1} \tag{9}$$

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

$$\mathbb{P}(T > c^{SN}(\alpha)) \leq \alpha + Cn_{-c_1} \tag{10}$$

Moreover, this bound holds uniformly over all distributions $\mathcal{L}_X$ satisfying the moment conditions as well as the above requirement (9). In addition, if (9) 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

$$\mathbb{P}(T > c^{SN}(\alpha)) \to 1 - e^{-\alpha}$$

I think the last bit is just to show that the test is approximately non-conservative.

Two-step method: Now move to combine the SN method with inequality selection. Motivation for doing this is that when $\mu_j < 0$ for some $j = 1, \dots, p$ the inequality in (6) becomes strict. So, when there are many j for which μ_j are negative and large in absolute value, the resulting rest with one-step SN critical values would tend to be unnecessarily conservative. So, 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.

Formally, let $0 < \beta_n < \alpha/2$ be some constant. For generality, allow β_n to depend on n. In particular, we allow $\beta_n = o(1)$. Let $c^{SN}(\beta_n)$ be the SN critical value with size β_n and define the set $\hat{J}_{SN} \subset \{1, \ldots, p\}$ by

$$\hat{J}_{SN} := \left\{ j \in \{1, \ldots, p\} : \sqrt{n}\hat{\mu}_j / \hat{\sigma}_j > -2c^{SN}(\beta_n) \right\} \tag{11}$$

Let $\hat{k} = |\hat{J}_{SN}|$. Then, the two step SN critical value is defined by

$$c^{SN,2S}(\alpha) = \begin{cases} \frac{\Phi^{-1}(1-(\alpha-2\beta_n)/\hat{k})}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta_n)/\hat{k})}}, & \text{if } \hat{k} \geq 1 \\ 0, & \text{if } \hat{k} = 0 \end{cases} \tag{12}$$

Then paper claims the following theorem

Theorem 3. *Suppose there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

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

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

$$\mathbb{P}(T > c^{SN,2S}(\alpha)) \leq \alpha + Cn^{-c} \tag{13}$$

Moreover, this bound holds uniformly over all distribution $\mathcal{L}_X$ satisfying (6) and the above condition. In addition, if all components of X_1 are independent, $\mu_j = 0$ and $p = p_n \to \infty$ while $\beta_n \to 0$ then

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

2.4.2 Bootstrap Methods

Section considers Multiplier Bootstrap and Empitical Bootstrap methods. These methods are computationally harder but they lead to less conservative tests.

One-Step Method First consider the one-step method (without moment selection). In order to make the test have size α, it is enough to choose the critical value as a bound on the $(1-\alpha)$ quantile of the distribution of

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

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

Alternatively, we consider a Gaussian approximation. Under suitable regularity conditions

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

where $Z_i = (Z_{i1}, \ldots, Z_{ip})^T$ are defined above ($Z_j = (X_j - \mu_j)/\sigma_j$). When p is fixed, the central limit theorem shows that, as $n \to \infty$,

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

By the continuous mapping theorem, this gives us that

$$\max_{1\le j\le p} \sqrt{n}\mathbb{E}_n[Z_{ij}] \rightsquigarrow \max_{1\le j\le p} Y_j$$

so we can take the critical value to be the $(1-\alpha)$ wuantile of $\max_{1\le j\le p} Y_j$. This theory does not cover when p grows with n. Different tools should be used to derive an appropriate critical value for the test. A 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. However, such Berry Esseen bounds require p to be small in comparasion with n in order to garuntee that the distribution of $\sqrt{n}\mathbb{E}_n Z_i$ is similar to that of Y. This approach builds on the work of (Chernozhukov, Chetverikov, and Kato, 2013, 2017) to show that, under some mild regularity conditions, the distribution of $\max_{1\le j\le p} \sqrt{n}\mathbb{E}_n[Z_{ij}]$ can be approximated by that of $\max_{1\le j\le p}$ in the sense of Kolmogrov distance even when p is larger or much larger than n.

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

Algorithm (Multiplier bootstrap)

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

$$W^{MB} = \max_{1\le j\le p} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \hat{\mu}_j)]}{\hat{\sigma}_j} \tag{14}$$

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

Algorithm (Empirical bootstrap)

1. Generate a bootstrap sample $X_1^*, \dots X_n^*$
2. Construct the empirical bootstrap test statistic

$$W^{EB} = \max_{1\le j\le p} \frac{\sqrt{n}\mathbb{E}_n[X_{ij}^* - \hat{\mu}_j]}{\hat{\sigma}_j} \tag{15}$$

3. Calculate $c^{EB}(\alpha)$ as the contional $(1-\alpha)$ quantile of W^{EB} given X_1^n.

We call these the one step multipler bootstrap and empirical bootsrap critical values, respectively, with size α. Can be computed with any precision using simulation.

Intuitively it is expected that the multiplier bootstrap works well since, conditional on the data, the vector

$$\left(\frac{\sqrt{n}\mathbb{E}[\epsilon_i(x_{ij} - \hat{\mu}_j)]}{\sigma_j} \right)_{1\le j\le p}$$

has the centered normal distribution with covariance matrix

$$\mathbb{E}_n \left[\frac{(X_{ij} - \hat{\mu}_j)}{\hat{\sigma}_j} \frac{(X_{ik} - \hat{\mu}_k)}{\hat{\sigma}_k} \right], 1 \le j, k \le p \tag{16}$$

which should be close to the covariance matrix of the vector Y. Indeed by Theorem 2 in Chenozhukov, Chetverikov, and Kato (2015), the primary factor for the bound on the Kolmogorov [4] distance between the

[4]The Kolmogorov Distance is defined as, for two pr. measures μ, ν on $\mathbb{R}$, $\mathrm{Kolm}(\mu,\nu) := \sup_{x\in\mathbb{R}} \left| \mu\left((-\infty, x]\right) - \nu\left((-\infty, x]\right) \right|$

conditional distribution of W and the distribution of $\max_{1\leq j\leq p} Y_j$ is

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

which is shown to be small even when $p \gg n$ (under suitable conditions).

Following theorem establishes validity of the MB and EB critical values.

Theorem 4 (Validity of one-step MB and EB methods)**.** *Let $c^B(\alpha)$ stand for either $c^{MB}(\alpha)$ or $c^{EB}(\alpha)$. Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

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

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

$$\mathbb{P}(T < c^B(\alpha)) \leq \alpha + Cn^{-c} \tag{18}$$

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

$$\left|\mathbb{P}(T > c^B(\alpha)) - \alpha\right| \leq Cn^{-c} \tag{19}$$

Moreover both bounds hold uniformly over all distributions L_X satisfying the conditions (4) and (17).

Leave analysis of more general exchangeable weighted bootstraps in the high dimensional setting for future works. Also observe that the condition (17) required fof the validity of the one-step MB/EB methods is stronger than qhat is required for validity of the two-step SN method.

Two-step Methods Now consider combining bootstrap methods with inequality selection. To describe, let $0 < \beta_n < \alpha/2$ be some constant. As before, β_n can 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, repsectively. Define the sets $\hat{J}_{MB}$ and $\hat{J}_{EB}$ by

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

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 $\epsilon_1,\ldots,\epsilon_n$ independent of the data X_1^n.
2. Construct the multiplier bootstrap test statistic

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

3. Calculate $c^{MB,2S}$ as the conditional $(1-\alpha+2\beta_n)$-quantile of $W_{\hat{J}_{MB}}$ given the data

Algorithm (Empirical bootstrap with inequality selection).

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

2. Construct the empirical bootstrap test statistic

$$W_{\hat{J}_{EB}} = \begin{cases} \max_{j \in \hat{J}_{EB}} \frac{\sqrt{n}\mathbb{E}_n[X^*_{ij} - \hat{\mu}_j]}{\hat{\sigma}_j} & \text{if } \hat{J}_{EB} \text{ is not empty} \\ 0 & \text{otherwise} \end{cases}$$

3. Calculate $c^{EB,2S}(\alpha)$ as the conditional $(1-\alpha+2\beta_n)$-quantile of $W_{\hat{J}_{EB}}$ given the data

Theorem 5 (Validity of two-step MB and EB methods)**.** *Let $c^{B,2S}(\alpha)$ stand for either $c^{MB,2S}(\alpha)$ or $c^{EB,2S}(\alpha)$. Suppose that the asusmption of Theorem 4 is satisfied. Moreover, suppose that $\log(1/\beta_n) \leq C_1 \log n$. Then there exist positive constants c, C depending only on c_1, C_1 such that under H_0,*

$$\mathbb{P}(T > c^{B,2S}(\alpha)) \leq \alpha + Cn^{-c}$$

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

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

so that under an extra assumption that $\beta \leq C_1 n^{-c_1}$[5]

$$\left|\mathbb{P}(T > c^{B,2S}(\alpha)) - \alpha\right| \leq Cn^{-c}$$

Moreover all these bounds hold uniformly over all distributions L_X satisfying (4) and (17)

It is sort of interesting to note that all these theorems are "non-asymptotic" in the sense that if the conditions hold then these inequalities "really" hold.

2.4.3 Hybrid Methods

Have considered one-step SN, MB, and EB methods and their two-step variatns. In fact, can also consider hybrids of these methods. For example, can use the SN method for inquality selection and then apply the MB or EB method for the selected inequalities, which is computationally more tractable. Notate this as the HB method. Formally, let $0 < \beta_n < \alpha/2$ be some constants and recall the set $\hat{J}_{SN} \subset \{1, \ldots, p\}$ defined above. Then the hybrid MB critical value, $c^{MB,H}(\alpha)$ is defined by the following procudure:

Algorithm (Multiplier Bootstrap Hybrid method).

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

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

3. Calvculate $c^{MB,H}(\alpha)$ as the contional $(1-\alpha+2\beta_n)$-quantile of $W_{\hat{J}_{SN}}$ given the data.

This can be equibalently defined for the empirical bootstrap.

Theorem 6 (Validity of hybrid two-step methods)**.** *Let $c^{MB,H}$ stand either for $c^{MB,H}(\alpha)$ or $c^{MB,H}(\alpha)$. Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that (17) is verified. Moreover, suppose that $\log(1/\beta_n) \leq C_1 \log n$. Then all the conclusions of Theorem 5 hold with $c^{B,MS}(\alpha)$ replaced by $c^{B,H}(\alpha)$.*

[5] which is to say β_n goes to 0 reasonable fast

2.4.4 Three-step method

In empirical studies based on moment inqualites one generally has inequalities of the form

$$\mathrm{E}[g_j(\xi,\theta)] \le 0 \qquad \text{for all } j = 1,\dots,p \tag{20}$$

where ξ is a vector of r.v's from a distribution denoted $\mathcal{L}_\xi$, $\theta = (\theta_1,\dots,\theta_r)^T$ 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) and (2) arise when one tests the null hypothesis $\theta = \theta_0$ against the alternative $\theta \ne \theta_0$ 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, have shown how to increase the poser of such tests by employing inequality selection procedures that allow the researcher to drop uninformative inequalities. In this subsection, combine this selection procedure with another procedure suitable for the model (20) 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 (20), 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, 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 only consider the case where the function $\theta \mapsto g_j(\xi,\theta)$ are almost everywhere continuously differentiable and leave the extension to non-differentiable functions to future work.

Start with the 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 and alternative hypothesis

$$\begin{aligned} H_0 &: \mathrm{E}[g_j(\xi,\theta_0)] \le 0 \quad \text{for all } j = 1,\dots,p \\ H_a &: \mathrm{E}[g_j(\xi,\theta_0)] > 0 \quad \text{for some } j = 1,\dots,p \end{aligned}$$

where θ_0 is some value of the parameter θ. Define

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

Further, let $X_{ij} := g_j(\xi_i,\theta_0), \mu_j := \mathrm{E}[X_{1j}], \sigma_j := \big(\mathrm{Var}(X_{1j})\big)^{1/2}, V_{ijl} := m_jl(\xi_i,\theta_0), \mu^B_{jl} = \mathrm{E}[V_{1jl}]$, and $\sigma^V_{jl} := \big(\mathrm{Var}(V_{1jl})\big)^{1/2}$. Assume that

$$\mathrm{E}[X^2_{1,j}] < \infty, \sigma_j > 0, j = 1,\dots,p \tag{21}$$

$$\mathrm{E}[X^2_{1,j,l}] < \infty, \sigma^V_{jl} > 0, j = 1,\dots,p, l = 1,\dots,r \tag{22}$$

In addition, let $\hat\mu_j = \mathbb{E}_n[X_{ij}]$ and $\hat\sigma_j = \left(\mathbb{E}\left[(X_{ij}-\hat\mu_j)^2\right]\right)^{1/2}$ be estimators of μ_j and σ_j, respectively. Similarly let $\hat\mu^V_{jl} = \mathbb{E}_n[V_{ijl}]$ and $\hat\sigma^V_{jl} = \left(\mathbb{E}\left[(V_{ij}-\hat\mu^V_{jl})^2\right]\right)^{1/2}$ be esttimators of μ^V_{jl}. The inequality selection derived is similar to the bootstrap methods described in Section 4

Algorithm(Multiplier bootstrap for gradient statistic).

1. Generate independent standard normal variables $\epsilon_1,\dots,\epsilon_n$ independent of the data.
2. Construct the multiplier bootstrap gradient statistic

$$W^V_{MB} = \max_{j,l} \frac{\sqrt{n}|\mathbb{E}_n[\epsilon_i(V_{ijl}-\hat\mu^V_{jl})]|}{\hat\sigma^V_{jl}} \tag{23}$$

3. For $\gamma \in (0,1)$, calculate $c^{MB,V}(\gamma)$ as the conditional $(1-\gamma)$ quantile of W^V_{MB} given the data.

Algorithm(Empirical bootstrap for gradient statistic).

1. Generate a bootstrap sample $V_1^*, \ldots, V_n^*$ as i.i.d draws from the data
2. Construct the empirical bootstrap gradient statistic

$$W^V_{EB} = \max_{j,l} \frac{\sqrt{n}|\mathbb{E}_n[V^*_{ijl} - \hat{\mu}^V_{jl}]|}{\hat{\sigma}^V_{jl}} \tag{24}$$

3. For $\gamma \in (0,1)$, calculate $c^{EB,V}(\gamma)$ as the conditional $(1-\gamma)$ quantile of W^V_{EB} given the data.

For $c_2, C_2 > 0$, 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 $\log\big(1/(\beta_n - \varphi_n)\big) \leq C_2 \log n$ where α is the nominal level of the test. Define three estimated sets of inequalities

$$\begin{aligned}
\hat{J}_B &:= \Big\{ j \in \{1, \ldots, p\} : \sqrt{n}\,\hat{\mu}_j \,/\, \hat{\sigma}_j > -2c^B(\beta_n) \Big\} \\
\hat{J}'_B &:= \Big\{ j \in \{1, \ldots, p\} : \sqrt{n}|\,\hat{\mu}^V_{jl} \,/\, \hat{\sigma}^V_{jl}\,| > 3c^{B,V}(\beta_n - \phi_n) \text{ for some } l = 1, \ldots, r \Big\} \\
\hat{J}''_B &:= \Big\{ j \in \{1, \ldots, p\} : \sqrt{n}|\,\hat{\mu}^V_{jl} \,/\, \hat{\sigma}^V_{jl}\,| > c^{B,V}(\beta_n + \phi_n) \text{ for some } l = 1, \ldots, r \Big\}
\end{aligned}$$

where B stands for either MB or EB.

The derived weakly informative inequality selection procedure requires that both the test statistic and the critical value depend on the estimated sets of inequalities. Let T^B and $c^{B,3S}$ denote the test statistic and the critical value. If the set $\hat{J}'_B$ is empty, set the test statistic and critical value $T^B = c^{B,3S} = 0$. Otherwise, define the test statistic

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

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 $\hat{J}_B$ replaced by $\hat{J}' \cap \hat{J}''_B$ and also $2\beta_n$ replaced by $4\beta_n$. That is $c^{B,3S}(\alpha)$ is the conditional $(1-\alpha+4\beta_n)$-quantile of $W_{\hat{J}'_B \cap \hat{J}''_B}$ given the data.

Stating the main results of this section requires the following notation. Let

$$Z^V_{ijl} := (V_{ijl} - \mu^V_{ijl})/\sigma^V_{jl} \quad \text{and} \quad M^V_{n,k} := \max_{j,l} \Big(\mathrm{E}[|Z^V_{ijl}|^k] \Big)^{1/k} \quad \text{and} \quad B^V_n := \Big(\mathrm{E}[\max_{j,l} (Z^V_{ijl})^4 \Big)^{1/4}$$

Theorem 7 (Validity of three-step MB and EB methods)**.** [6] *Let T^B and $c^{B,3S}(\alpha)$ stand for T^{MB} and $c^{MB,3S}(\alpha)$ or for T^{EB} and $c^{EB,3S}(\alpha)$. Suppose there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$\Big(M^3_{n,3} \vee M^2_{n,4} \vee B_n)^2 \log^{7/2}(pn) \Big) \leq C_1 n^{1/2 - c_1} \tag{25}$$

and

$$\Big((M^V_{n,3})^3 \vee (M^V_{n,4})^2 \vee (B^V_n)^2 \log^{7/2}(pn) \Big) \leq C_1 n^{1/2 - c_1} \tag{26}$$

Moreover, suppose that $\log(1/(\beta_n - \phi_n)) \leq C_2 \log n$ and $\phi_n \log n \geq c_2$ for some constants $c_2, C_2 > 0$. Then there exist positive constants c, C depending only on c_1, C_1, c_2, C_2such that, under H_0,

$$\mathbb{P}(T^B > c^{B,3S}(\alpha)) \leq \alpha + Cn^{-c}$$

and this bound holds uniformly over all distributions L_ξ satisfying (21),(22),(25), and (26).

[6] If I understand correctly, this method only selects out weakly uninformative inequalites based on the gradient, not necessarily inequalities with say, $\hat{\mu}_j \ll 0$.

2.5 Power

Consider the same general setup as described in the intruction and assume that (4) holds. Pick any $\alpha \in (0, 1/2)$ and consider the test of the form

$$T > \hat{c}(\alpha) \implies \text{ reject } H_0$$

Where $\hat{c}(\alpha)$ 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 $c^{EB,H}(\alpha)$.[7] The following result holds:

Theorem 8 (Rate of uniform consistency)**.** *Suppose there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$M_{n,4}^2 \log^{1/2} p \leq C_1 n^{1/2 - c_1} \text{ and } \log^{3/2} p \leq C_1 n \tag{27}$$

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[8]

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

we have

$$\mathbb{P}(T > \hat{c}(\alpha)) \geq 1 - \frac{C}{\epsilon^2 \log(p/\alpha)} - C n^{-c}$$

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

$$\inf_{\mu \in \mathcal{B}_n} \mathbb{P}_\mu(T > \hat{c}(\alpha)) \geq 1 - o(1) \tag{28}$$

where

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

and P_μ denotes the probability measure for the distribution $\mathcal{L}_X$ having mean μ. Moreover, the above asymptotic result (28) holds uniformly with respect to any sequnce of distributions $\mathcal{L}_X$ satisfying (4) and (27).

Theorem 8 shows that tests are uniformly consistent against all alternatives excluding those in a small neighborhood of alternatives that are too close to the null. Size of this neighborhood is shrikning at a fast rate. They show in a working paper that no test can be consistent against all alternatives whose distance from the null converges to zero faster than $\sqrt{(\log p_n)/n}$ and that the tests above are minimax optimal.

2.6 Monte Carlo Experiments

Results are given, I would check document.

[7] Importantly, the three step procedure is not included here.
[8] Seems similar to LASSO condition