A Two-Step Method for Testing Many Moment Inequalities*

Yuehao Bai
Department of Economics
University of Michigan
yuehaob@umich.edu

Andres Santos
Department of Economics
University of California–Los Angeles
andres@econ.ucla.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

January 16, 2021

Abstract

This paper considers the problem of testing a finite number of moment inequalities. For this problem, Romano et al. (2014) propose a two-step testing procedure. In the first step, the procedure incorporates information about the location of moments using a confidence region. In the second step, the procedure accounts for the use of the confidence region in the first step by adjusting the significance level of the test appropriately. Its justification, however, has so far been limited to settings in which the number of moments is fixed with the sample size. In this paper, we provide weak assumptions under which the same procedure remains valid even in settings in which there are "many" moments in the sense that the number of moments grows rapidly with the sample size. We confirm the practical relevance of our theoretical guarantees in a simulation study. We additionally provide both numerical and theoretical evidence that the procedure compares favorably with the method proposed by Chernozhukov et al. (2019), which has also been shown to be valid in such settings.

KEYWORDS: High-dimensional inference, partial identification, bootstrap, moment inequalities, multi-sided hypothesis

JEL classification codes: C12, C14

*We thank Roy Allen and Denis Chetverikov for helpful comments. The research of the third author is supported by NSF Grant SES-1530661.

1 Introduction

Let $X_i, i = 1, \ldots, n$ be an independent and identically distributed (i.i.d.) sequence of random variables with distribution $P \in \mathbf{P}_n$ on $\mathbf{R}^{p_n}$ and consider the problem of testing

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

where

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

and $\mathbf{P}_{1,n} = \mathbf{P}_n \setminus \mathbf{P}_{0,n}$. Here, the inequality in (2) is intended to be interpreted component-wise and $\mathbf{P}_n$ is a "large" class of possible distributions for the observed data. By indexing both the number of moments, p_n, and the class of possible distributions, $\mathbf{P}_n$, by the sample size n, we anticipate asymptotic results that allow the number of moments p_n to grow rapidly with the sample size n. In this way, our asymptotic framework can accommodate settings in which it is desired to test possibly "many" moment inequalities. Our goal is to construct tests $\phi_n = \phi_n(X_1, \ldots, X_n)$ of (1) that are uniformly consistent in level, i.e.,

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

for some pre-specified value of $\alpha \in (0, \frac{1}{2})$.

In many instances where the testing problem described above arises in economics, the number of moments is large. Examples include entry models, as in Ciliberto and Tamer (2009), in which p_n is on the order of 2^{m+1}, where m is the number of firms, and dynamic models of imperfect competition, as in Bajari et al. (2007), where p_n may even be as large as 500. Yet, with the notable exception of Chernozhukov et al. (2019), tests of (1) that have been proposed have only been shown to satisfy (3) under restrictions on $\mathbf{P}_n$ that require number of moments p_n to be small in the sense that it is independent of the sample size n. Canay and Shaikh (2017) provide a detailed review of these tests. In this paper, we focus on one particular such test of (1): the two-step testing procedure proposed by Romano et al. (2014). This test was shown to satisfy (3) under assumptions on $\mathbf{P}_n$ that restrict p_n to not depend on n. Romano et al. (2014) emphasize, however, that the test remains computationally feasible even if the number of moments is large, thereby permitting its implementation in examples such as those described above. In this paper, we show that the test of Romano et al. (2014) in fact continues to satisfy (3) for a large class of distributions that permits the number of moments p_n to grow exponentially with the sample size n. In this way, our results establish the validity of the methodology for testing "many" moment inequalities, thereby supporting its application in examples such as those described above.

Our theoretical analysis relies crucially on the seminal work of Chernozhukov et al. (2013, 2017) on the high-dimensional central limit theorem. The high-dimensional central limit theorem had previously been applied to study tests of (1) by Chernozhukov et al. (2019), who, as mentioned previously, develop tests that satisfy (3) for a large class of distributions $\mathbf{P}_n$ that permits the number of moments p_n to grow rapidly with the sample size n. One motivation for establishing that the test of Romano et al. (2014) remains valid with "many" moments is a result by Allen (2018) that provides conditions under which the test of Romano et al. (2014) *always* rejects whenever the preferred test in Chernozhukov et al. (2019) rejects. In Section 2.1, we revisit the arguments in Allen (2018) and highlight that the power advantages of Romano et al. (2014) arise from its use of a better bound for the nuisance parameter $\sqrt{n}E_P[X_i]$ than that employed by Chernozhukov et al. (2019). Prior to the results in this paper, however, it was unclear whether it was sensible to compare the power of tests developed by Chernozhukov et al. (2019) with the one proposed by Romano et al. (2014) because it was not known whether the latter test continued to satisfy (3) when the number of moments p_n was permitted to grow rapidly with the sample size n. In light of the results in this paper, such a comparison is now theoretically justified. In particular, we note that the power advantages established by Allen (2018) and the minimax rate optimality of the tests in Chernozhukov et al. (2019) imply that, under suitable conditions, the test in Romano et al. (2014) is also minimax rate optimal; see Remark 2.3 below. Since the result by Allen (2018) pertains a particular implementation of the test in Romano et al. (2014), we supplement our theoretical comparison with simulation evidence for other implementations of the two tests. In our simulations, we find that the test proposed by Romano et al. (2014) continues to compare favorably, both in terms of size and power, with the test proposed by Chernozhukov et al. (2019).

The remainder of the paper is organized as follows. In Section 2, we provide a detailed description of the testing procedure in Romano et al. (2014) and the assumptions that will underlie our analysis. In our discussion of the assumptions, we emphasize that they permit the number of moments p_n to grow rapidly with the sample size n. We then establish that the test satisfies (3) under these assumptions. The proof of this result is relegated to the Appendix. In Section 2, we also revisit the analysis in Allen (2018) to better understand the power advantages of the test proposed by Romano et al. (2014). Finally, in Section 3 we examine the practical relevance of our theoretical results via a simulation study, which includes further comparisons with the test proposed by Chernozhukov et al. (2019).

2 Main Result

We begin this section by describing the testing procedure in Romano et al. (2014). In order to do so, it is useful to introduce some further notation. For $1 \le j \le p_n$, let $X_{i,j}$ denote the jth component of X_i and set

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

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

We will also make use of the notation $\mu_j(P) \equiv E_P[X_{i,j}]$ and $\sigma^2_j(P) \equiv \mathrm{Var}_P[X_{i,j}]$, so (4) may be equivalently expressed as $\mu_j(\hat{P}_n)$ and (5) as $\sigma^2_j(\hat{P}_n)$, where $\hat{P}_n$ is the empirical distribution of $\{X_i\}_{i=1}^n$. While Romano et al. (2014) consider a variety of test statistics, we focus on the test that rejects for large values of

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

In order to define the critical value with which we will compare T_n, it will be 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 by $X_i^*, i = 1, \ldots, n$. We further define $\bar{X}_{j,n}^*$ and $(S_{j,n}^*)^2$ by analogy with $\bar{X}_{j,n}$ in (4) and $S_{j,n}^2$ in (5) but substituting X_i^* for X_i. Using this notation, the critical value with which we will compare T_n is given by

$$\hat{c}_n^{(2)}(1-\alpha+\beta) \equiv \inf\left\{c \in \mathbf{R} : P\left\{\max\left\{\max_{1\leq j\leq p_n} \frac{\sqrt{n}(\bar{X}_{j,n}^* - \bar{X}_{j,n} + \hat{u}_{j,n})}{S_{j,n}^*}, 0\right\} \leq c \Big| \{X_i\}_{i=1}^n\right\} \geq 1-\alpha+\beta\right\} , \quad (6)$$

where $\alpha \in (0, \frac{1}{2})$ is the nominal level of the test, $0 < \beta < \alpha$, and

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

with

$$\hat{c}_n^{(1)}(1-\beta) \equiv \inf\left\{c \in \mathbf{R} : P\left\{\max_{1\leq j\leq p_n} \frac{\sqrt{n}(\bar{X}_{j,n} - \bar{X}_{j,n}^*)}{S_{j,n}^*} \leq c \Big| \{X_i\}_{i=1}^n\right\} \geq 1-\beta\right\} . \quad (8)$$

The test ϕ_n^{RSW} of the null hypothesis in (1) we consider rejects whenever T_n exceeds $\hat{c}_n^{(2)}(1-\alpha+\beta)$, i.e.,

$$\phi_n^{\mathrm{RSW}} \equiv I\left\{T_n > \hat{c}_n^{(2)}(1-\alpha+\beta)\right\} . \quad (9)$$

In order to motivate this choice of critical value, it is useful to note the test statistic T_n satisfies

$$T_n = \max\left\{\max_{1\leq j\leq p_n}\left(\frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P))}{S_{j,n}} + \frac{\sqrt{n}\mu_j(P)}{S_{j,n}}\right), 0\right\} . \quad (10)$$

The decomposition of T_n in (10) highlights that the main impediment to approximating the distribution of T_n is the presence of the nuisance parameters $\sqrt{n}\mu_j(P)$ for $1 \leq j \leq p_n$. Even 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 4.1 in the Appendix employs their insight and the high-dimensional central limit theorem of Chernozhukov et al. (2017) to show, under conditions that permit p_n to grow rapidly with the sample size n, that $\sqrt{n}\mu_j(P) \leq \sqrt{n}\hat{u}_{j,n}$ for all $1 \leq j \leq p_n$ with probability 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 function of these nuisance parameters, any quantile of T_n is maximized over said confidence region by setting $\sqrt{n}\mu_j(P) = \sqrt{n}\hat{u}_{j,n}$ for all $1 \leq j \leq p_n$. Thus, 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 $\sqrt{n}\mu_j(P) = \sqrt{n}\hat{u}_{j,n}$ for all $1 \leq j \leq p_n$. Here, the $1-\alpha+\beta$ quantile is employed instead of $1-\alpha$, to account for the possibility that, with probability approximately no greater than β, we may find $\sqrt{n}\mu_j(P) > \sqrt{n}\hat{u}_{j,n}$ for some $1 \leq j \leq p_n$.

Remark 2.1. Instead of testing (1), in certain applications it is of interest to test whether P satisfies

$$\mu_j(P) = 0 \text{ for all } 1 \leq j \leq k_n \text{ and } \mu_j(P) \leq 0 \text{ for all } k_n + 1 \leq j \leq p_n.$$

While such a hypothesis can be mapped into our framework simply by writing $\mu_j(P) = 0$ as the inequalities $\mu_j(P) \leq 0$ and $-\mu_j(P) \leq 0$, a direct application of the test ϕ_n^{RSW} is not advisable because it does not take full advantage of the structure of the null hypothesis. Formally, constructing a confidence region for $\sqrt{n}\mu_j(P)$ for all $1 \leq j \leq p_n$ is not needed as we now know that, under the null hypothesis, $\sqrt{n}\mu_j(P) = 0$ for all $1 \leq j \leq k_n$. As a result, Romano et al. (2014) instead advocate employing the test statistic

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

substituting the maximum over $1 \leq j \leq p_n$ with a maximum over $k_n + 1 \leq j \leq p_n$ when computing $\hat{c}_n^{(1)}(1-\beta)$, and setting the $1 - \alpha + \beta$ (conditional on $\{X_i\}_{i=1}^n$) bootstrap quantile of the statistic

$$\max\left\{\max_{1\leq j\leq k_n}\left|\frac{\sqrt{n}(\bar{X}^*_{j,n} - \bar{X}_{j,n})}{S^*_{j,n}}\right|, \max_{1\leq j\leq p_n}\frac{\sqrt{n}(\bar{X}^*_{j,n} - \bar{X}_{j,n} + \hat{u}_{j,n})}{S^*_{j,n}}, 0\right\}$$

as the critical value with which to compare T_n; see Remarks 2.3 and S.4 in Romano et al. (2014). ■

Our analysis of the test defined in (9) requires the following assumption:

Assumption 2.1. (i) $\{X_i\}_{i=1}^n$ is an i.i.d. sample with $X_i \in \mathbf{R}^{p_n}$ and $X_i \sim P \in \mathbf{P}_n$; (ii) $\sigma_j(P) > 0$ for all $1 \leq j \leq p_n$ and $P \in \mathbf{P}_n$; (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 $E_P[\max_{1\leq j\leq p_n}|X_{i,j} - \mu_j(P)|^4/\sigma_j^4(P)] \leq B_n^4$ 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 $\delta \in (0, 1)$.

Assumption 2.1(i) simply formalizes the requirement that $\{X_i\}_{i=1}^n$ be an i.i.d. sample, while Assumption 2.1(ii) requires the variance of $X_{i,j}$ to be positive for all $P \in \mathbf{P}_n$ and $1 \leq j \leq p_n$. In Assumption 2.1(iii), we impose a uniform in $P \in \mathbf{P}_n$ and $1 \leq j \leq p_n$ bound on the (standardized) moments of $X_{i,j}$. This condition is a strengthening of the (standardized) uniform integrability condition imposed by Romano et al. (2014), which we require in order to study a setting in which p_n diverges to infinity. Assumption 2.1(iv) bounds the fourth moments of the maximum of the (standardized) $X_{i,j}$. If, for example, the support of the standardized $X_{i,j}$ under P is bounded uniformly in $P \in \mathbf{P}_n$, $1 \leq j \leq p_n$, and n, then B_n can be taken to be a constant independent of n. In contrast, if the standardized $X_{i,j}$ have exponential tails uniformly in $P \in \mathbf{P}_n$, $1 \leq j \leq p_n$, and n, then B_n can be set proportional to a power of $\log(p_n)$. Finally, Assumption 2.1(v) states the main condition governing the relationship between the dimension p_n and the sample size n. Importantly, we note that under suitable moment restrictions on $X_{i,j}$, p_n may grow exponentially with n.

Under Assumption 2.1, we are able to establish the main result of this paper.

Theorem 2.1. *If Assumption 2.1 holds, $\alpha \in (0, \frac{1}{2})$, and $0 < \beta < \alpha$, then ϕ_n^{RSW} defined in (9) satisfies (3).*

Theorem 2.1 verifies that the test proposed in Romano et al. (2014) is indeed able to satisfy (3) even in settings in which p_n grows rapidly with the sample size. In this manner, Theorem 2.1 provides theoretical support for applying the test ϕ_n^{RSW} is empirical applications with "many" moment inequalities. The ability of the test in Romano et al. (2014) to control size in such high-dimensional settings had previously been conjectured, but not established, by Chernozhukov et al. (2019).

While Theorem 2.1 applies for any fixed value of $\beta \in (0, \alpha)$, we note that the theorem remains true if β is instead allowed to depend on n provided $\beta_n \in (0, \alpha)$ for all n (but with β_n possibly converging to $\{0, \alpha\}$). Such an extension can be helpful, for example, when a researcher has a set of local alternatives against which she aims to maximize (over β) weighted average power; see Remark S.6 in Romano et al. (2014). In such a setting, the optimal β can depend on n through the dependence of p_n on n. We emphasize, however, that the "optimal" β depends on the set of local alternatives under consideration. As a simple rule of thumb, we find that setting $\beta = \alpha/10$, as recommended by Romano et al. (2014), performs well in our simulations.

Remark 2.2. In some cases, it may be of interest to determine not just whether $\mu_j(P) \leq 0$ for all $1 \leq j \leq p_n$ or not, but the specific values of $1 \leq j \leq p_n$ for which $\mu_j(P) > 0$. For this purpose, it is natural to consider the problem of simultaneously testing $H_j : P \in \mathbf{P}_{j,n}$ versus $H_j' : P \in \mathbf{P}_{j,n}'$ for $j = 1, \ldots, p_n$, where $\mathbf{P}_{j,n} \equiv \{P \in \mathbf{P}_n : \mu_j(P) \leq 0\}$ and $\mathbf{P}_{j,n}' \equiv \mathbf{P}_n \setminus \mathbf{P}_j$. In order to account for the multiplicity of decisions being made, it is common to require control of the familywise error rate in the sense that

$$\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_n} FWER_P \leq \alpha \ , \tag{11}$$

where

$$FWER_P = P\{\text{reject any } H_j \text{ with } P \in \mathbf{P}_{j,n}\} \ .$$

Using Theorem 2.1, it is possible to develop procedures that satisfy (11) under Assumption 2.1. For instance, it is straightforward to show that the procedure that rejects any H_j with $\sqrt{n}\bar{X}_{j,n}/S_{j,n} > \hat{c}_n^{(2)}(1 - \alpha + \beta)$ satisfies (11) under Assumption 2.1. By combining Theorem 2.1 with results in Romano and Wolf (2005), iterative improvements upon such a procedure are also possible. Indeed, one may simply apply this procedure and then repeat it with the set of null hypotheses that are *not* rejected after the first application, continuing in this fashion until no further null hypotheses are rejected. For some results in settings in which p_n remains fixed with the sample size n, see Romano and Wolf (2018). ■

2.1 Alternative Procedures

Chernozhukov et al. (2019) propose several different tests of (1). In our comparisons, we restrict attention to their most preferred test, which is similar in spirit to the "generalized moment selection" tests developed in Andrews and Soares (2010). The proposed test rejects for large values of

$$\tilde{T}_n \equiv \max_{1 \leq j \leq p_n} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} \ .$$

In order to describe the critical value with which they compare $\tilde{T}_n$, for $I \subseteq \{1, \ldots, p_n\}$ and $\gamma \in (\frac{1}{2}, 1)$, define

$$\tilde{c}_n(I, \gamma) \equiv \inf \left\{ c \in \mathbf{R} : P\left\{ \max_{j \in I} \frac{\sqrt{n}(\bar{X}_{j,n}^* - \bar{X}_{j,n})}{S_{j,n}^*} \leq c \Big| \{X_i\}_{i=1}^n \right\} \geq \gamma \right\} \ . \tag{12}$$

Using this notation, the proposed test ϕ_n^{CCK} rejects whenever $\tilde{T}_n$ exceeds $\tilde{c}_n(\hat{I}_n, 1-\alpha+2\beta)$, where

$$\hat{I}_n \equiv \left\{1 \le j \le p_n : \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} > -2\tilde{c}_n(\{1,\dots,p_n\}, 1-\beta)\right\} ,$$

$\alpha \in (0, \frac{1}{2})$ and $0 < \beta < \frac{\alpha}{2}$, i.e.,

$$\phi_n^{\text{CCK}} \equiv I\{\tilde{T}_n > \tilde{c}_n(\hat{I}_n, 1-\alpha+2\beta)\} . \tag{13}$$

In our simulations, we also consider the test ϕ_n^{CCK2} defined as above, but in which $S_{j,n}^*$ in (12) is replaced with $S_{j,n}$. It is worth emphasizing that the formal analysis in Chernozhukov et al. (2019) concerns ϕ_n^{CCK2}, but we include both tests in our simulations for completeness.

Allen (2018) showed that a version of the test in Romano et al. (2014) is more powerful than the preferred test in Chernozhukov et al. (2019) in the sense that the former *always* rejects the null hypothesis whenever the latter rejects the null hypothesis. An inspection of the proof of Allen (2018) reveals that ϕ_n^{RSW} is more powerful than ϕ_n^{CCK} in the sense that $\phi_n^{\text{RSW}} \ge \phi_n^{\text{CCK}}$ (with probability one) if one employs a Gaussian multiplier bootstrap instead of the empirical bootstrap. Similarly, it is also possible to show that a version of ϕ_n^{RSW} that replaces $S_{j,n}^*$ in (6) and (7) with $S_{j,n}$, which we denote by ϕ_n^{RSW2}, satisfies $\phi_n^{\text{RSW2}} \ge \phi_n^{\text{CCK2}}$ (with probability one) provided that a Gaussian multiplier bootstrap is used instead of the empirical bootstrap.

In order to gain some intuition for the power advantage of Romano et al. (2014) it is helpful to revisit the arguments behind Allen (2018) in a stylized Gaussian model. Specifically, suppose that $X \sim N(\mu, \Sigma)$ with unknown mean $\mu \in \mathbf{R}^p$ and known $p \times p$ covariance matrix Σ. In this setting there is no need to bootstrap and when implementing ϕ_n^{CCK} we can replace $\tilde{c}_n(I, \gamma)$ (as defined in (12)) by the quantile

$$\tilde{c}_n^{\mathrm{g}}(I, \gamma) \equiv \inf\left\{c \in \mathbf{R} : P\left\{\max_{j \in I} \frac{Z_j}{\sigma_j} \le c\right\} \ge \gamma\right\}$$

where $Z \sim N(0, \Sigma)$. Further setting $\hat{I}_n^{\mathrm{g}} \equiv \{1 \le j \le p : \sqrt{n}\bar{X}_{j,n} > -2\sigma_j \tilde{c}_n^{\mathrm{g}}(\{1,\dots,p\}, 1-\beta)\}$ and observing that ϕ_n^{CCK} will not reject when $\tilde{T}_n < 0$ it follows from $T_n = \max\{\tilde{T}_n, 0\}$ that in this context we have

$$\phi_n^{\text{CCK}} = I\left\{T_n > \tilde{c}_n^{\mathrm{g}}(\hat{I}_n^{\mathrm{g}}, 1-\alpha+2\beta)\right\}.$$

Analogously, when implementing ϕ_n^{RSW} we may replace $\hat{c}_n^{(1)}(1-\beta)$ (as defined in (8)) by $\tilde{c}_n^{\mathrm{g}}(\{1,\dots,p\}, 1-\beta)$ and $\hat{c}_n^{(2)}(1-\alpha+\beta)$ (as defined in (6)) by the quantile

$$c_n^{\mathrm{g}}(1-\alpha+\beta) \equiv \inf\left\{c \in \mathbf{R} : P\left\{\max\left\{\max_{1 \le j \le p_n} \frac{Z_j}{\sigma_j} + \frac{\sqrt{n}\hat{u}_{j,n}^{\mathrm{g}}}{\sigma_j}, 0\right\} \le c \,\middle|\, \{\hat{u}_{j,n}^{\mathrm{g}}\}_{j=1}^p\right\} \ge 1-\alpha+\beta\right\}$$

where

$$\hat{u}_{j,n}^{\mathrm{g}} \equiv \min\left\{\bar{X}_{j,n} + \frac{\sigma_j}{\sqrt{n}}\tilde{c}_n^{\mathrm{g}}(\{1,\dots,p\}, 1-\beta), 0\right\}.$$

Since $\phi_n^{\text{CCK}} = \phi_n^{\text{RSW}} = 0$ when $\hat{I}_n^{\mathrm{g}} = \emptyset$ and both tests are based on the statistic T_n, establishing $\phi_n^{\text{CCK}} \le \phi_n^{\text{RSW}}$

is equivalent to showing $\tilde{c}_n^{\mathrm{g}}(\hat{I}_n^{\mathrm{g}}, 1-\alpha+2\beta) \geq c_n^{\mathrm{g}}(1-\alpha+\beta)$ whenever $\hat{I}_n^{\mathrm{g}} \neq \emptyset$. To this end note

$$
\begin{aligned}
& P\left\{\max\left\{\max_{1\leq j\leq p_n} \frac{Z_j}{\sigma_j} + \frac{\sqrt{n}\hat{u}_{j,n}^{\mathrm{g}}}{\sigma_j}, 0\right\} > \tilde{c}_n^{\mathrm{g}}(\hat{I}_n^{\mathrm{g}}, 1-\alpha+2\beta)\right\} \\
& \leq P\left\{\max_{j\in\hat{I}_n^{\mathrm{g}}} \frac{Z_j}{\sigma_j} + \frac{\sqrt{n}\hat{u}_{j,n}^{\mathrm{g}}}{\sigma_j} > \tilde{c}_n^{\mathrm{g}}(\hat{I}_n^{\mathrm{g}}, 1-\alpha+2\beta)\right\} + P\left\{\max_{j\in\{1,\ldots,p\}\setminus\hat{I}_n^{\mathrm{g}}} \frac{Z_j}{\sigma_j} + \frac{\sqrt{n}\hat{u}_{j,n}^{\mathrm{g}}}{\sigma_j} > \tilde{c}_n^{\mathrm{g}}(\hat{I}_n^{\mathrm{g}}, 1-\alpha+2\beta)\right\} \\
& \leq P\left\{\max_{j\in\hat{I}_n^{\mathrm{g}}} \frac{Z_j}{\sigma_j} > \tilde{c}_n^{\mathrm{g}}(\hat{I}_n^{\mathrm{g}}, 1-\alpha+2\beta)\right\} + P\left\{\max_{j\in\{1,\ldots,p\}\setminus\hat{I}_n^{\mathrm{g}}} \frac{Z_j}{\sigma_j} > \tilde{c}_n^{\mathrm{g}}(\{1,\ldots,p\}, 1-\beta)\right\} \\
& \leq (\alpha - 2\beta) + \beta \qquad (14)
\end{aligned}
$$

where: (i) the first inequality follows from the union bound and $\tilde{c}_n^{\mathrm{g}}(\hat{I}_n^{\mathrm{g}}, 1-\alpha+2\beta) > 0$; (ii) the second inequality follows from $\hat{u}_{n,j} \leq 0$ for all $j \in \hat{I}_n^{\mathrm{g}}$ and $\sqrt{n}\hat{u}_{n,j} \leq -\tilde{c}_n^{\mathrm{g}}(\{1,\ldots,p\}, 1-\beta)$ for all $j \in \{1,\ldots,p\} \setminus \hat{I}_n^{\mathrm{g}}$; and (iii) the final inequality follows by the set inclusion $\{1,\ldots,p\} \setminus \hat{I}_n^{\mathrm{g}} \subseteq \{1,\ldots,p\}$. Thus, by definition of $c_n^{\mathrm{g}}(1-\alpha+\beta)$, result (14) implies $\tilde{c}_n^{\mathrm{g}}(\hat{I}_n^{\mathrm{g}}, 1-\alpha+2\beta) \geq c_n^{\mathrm{g}}(1-\alpha+\beta)$ and hence that $\phi_n^{\mathrm{CCK}} \leq \phi_n^{\mathrm{RSW}}$ as claimed.

The arguments in Allen (2018) further provide some intuition as to the circumstances under which we should expect ϕ_n^{RSW} to be strictly more powerful than ϕ_n^{CCK}. Specifically, we highlight:

1. On the set $\hat{I}_n$ of selected moments, the bootstrap approximation employed in ϕ_n^{CCK} replaces $\sqrt{n}\mu_j(P)$ by 0, while the bootstrap approximation employed in ϕ_n^{RSW} replaces $\sqrt{n}\mu_j(P)$ by $\sqrt{n}\hat{u}_{n,j} \leq 0$. For alternatives $\mu(P)$ such that $\mu_j(P)$ is "small" in absolute value and negative for some $1 \leq j \leq p$, we would expect $\sqrt{n}\hat{u}_{n,j}$ to be strictly negative on $\hat{I}_n$ with positive probability, leading to $\phi_n^{\mathrm{CCK}} < \phi_n^{\mathrm{RSW}}$ with positive probability – i.e., the second equality in (14) would hold strictly (provided $\beta > 0$).

2. On the set $\{1,\ldots,p\} \setminus \hat{I}_n$ of unselected moments, the bootstrap approximation employed in ϕ_n^{CCK} replaces $\sqrt{n}\mu_j(P)$ by $-\infty$. In order for ϕ_n^{CCK} to have correct size in instances in which $\sqrt{n}\mu_j(P)$ is "small" in absolute value but incorrectly set to $-\infty$, ϕ_n^{CCK} employs a $1-\alpha+2\beta$ quantile as a critical value. In contrast, the bootstrap approximation employed in ϕ_n^{RSW} replaces $\sqrt{n}\mu_j(P)$ by $\sqrt{n}\hat{u}_{n,j}$, which remains valid with probability $1-\beta$ even when $j \in \{1,\ldots,p\} \setminus \hat{I}_n$. This distinction causes a power difference that we expect to be increasing in β – i.e., increasing β makes the final inequality in (14) more likely to hold strictly due to $\{1,\ldots,p\} \setminus \hat{I}_n$ being more likely to be a "small" subset of $\{1,\ldots,p\}$. Selecting a large β is preferable for alternatives $\mu(P)$ for which $\sqrt{n}\mu_j(P)$ is "large" in absolute value and negative for some $1 \leq j \leq p$, and hence we expect this power difference to be important in those contexts.

Remark 2.3. In a working paper version (arXiv:1312.7614.v4), Chernozhukov et al. (2019) show that their tests are asymptotically minimax rate optimal when considering alternatives $P \in \mathbf{P}_{1,n}$ satisfying $\max_{1\leq j\leq p} \mu_j(P)/\sigma_j(P) \geq r_n$ for a sequence r_n. Combining such a result with the arguments in Allen (2018) who provides conditions under which $\phi_n^{\mathrm{CCK}} \leq \phi_n^{\mathrm{RSW}}$ imply that the tests we consider inherit the minimax rate optimality results established by Chernozhukov et al. (2019). ■

3 Simulations

In this section, we examine the finite-sample behavior of the test of (1) described in Section 2 via a small simulation study. We also compare its behavior with tests described in Section 2.1.

We begin by describing the distribution of X_i. Following Chernozhukov et al. (2019), we specify that

$$X_{i,j} = \theta(I\{1 \leq j \leq 0.05p_n\} + \varepsilon_{i,j}) - bI\{0.1p_n < j \leq p_n\} + \varepsilon_{i,j}$$

for $1 \leq i \leq n$ and $1 \leq j \leq p_n$, where $\varepsilon_i, i = 1, \ldots, n$ are i.i.d. with distribution $N(0, \Sigma)$. We consider four different models, which differ according to the values of b and Σ.

Model 1: $b = 0$, $\Sigma_{j,k} = 1$ for $1 \leq j, k \leq p_n$ with $j = k$ and ρ otherwise.

Model 2: $b = 0.8$, $\Sigma_{j,k} = 1$ for $1 \leq j, k \leq p_n$ with $j = k$ and ρ otherwise.

Model 3: $b = 0$, $\Sigma_{j,k} = \rho^{|j-k|}$ for $1 \leq j, k \leq p_n$.

Model 4: $b = 0.8$, $\Sigma_{j,k} = \rho^{|j-k|}$ for $1 \leq j, k \leq p_n$.

In Chernozhukov et al. (2019), Models 1 and 2 are referred to as "equicorrelated" and Models 3 and 4 as "autocorrelated." For each model, we consider the following different values of ρ, p_n and θ: $\rho \in \{0, 0.5, 0.9\}$, $p_n \in \{40, 100, 200\}$, and $\theta \in \{0, 0.2\}$. In all designs, the sample size n is set to equal one hundred, and all tests are implemented at a $\alpha = 0.05$ nominal level. We do not consider non-Gaussian errors or larger sample sizes here because our interest lies mainly in examining the ability of the different tests to exploit components of $E_P[X_i]$ that are strictly negative to increase power rather than other aspects of the asymptotic approximations, which should be common across all of the tests we consider. In all of our specifications

$$E_P[X_{i,j}] = \begin{cases} \theta & \text{if } 1 \leq j \leq 0.05p_n \\ -b & \text{if } 0.1p_n < j \leq p_n \\ 0 & \text{otherwise} \end{cases},$$

so the number of negative components of $E_P[X_{i,j}]$ is governed by whether $b = 0$ or 0.8. Finally, we observe that the null hypothesis is true when $\theta = 0$ and the alternative hypothesis is true when $\theta = 0.2$. In this way, our designs permit us to study both the size and power of the tests under consideration.

In our simulations below, we consider three different tests:

RSW: The test ϕ_n^{RSW} defined in (9).

RSW2: The test ϕ_n^{RSW2} described in Section 2.1.

CCK: The test ϕ_n^{CCK} defined in (13).

CCK2: The test ϕ_n^{CCK2} described in Section 2.1.

Recall that the only distinction between ϕ_n^{RSW} and ϕ_n^{RSW2} is that the former employs $S_{j,n}^*$ in the bootstrap samples, while the latter employs $S_{j,n}$. The same distinction differentiates ϕ_n^{CCK} and ϕ_n^{CCK2}. Following recommendations in Romano et al. (2014) and Chernozhukov et al. (2019), we first choose $\beta = 0.005$ when implementing ϕ_n^{RSW} and $\beta = 0.001$ when implementing ϕ_n^{CCK} and ϕ_n^{CCK2}. After discussing these results, we examine the extent to which the comparisons are robust to different choices of β for each test.

The results of our simulations are presented in Table 1. Columns labeled 'RSW', 'RSW2', 'CCK' and 'CCK2' display rejection probabilities (in percentage points) for the corresponding test. Columns labeled '≥CCK' and '≥CCK2' display, respectively, the percentage of replications where $\phi_n^{\text{RSW}} \geq \phi_n^{\text{CCK}}$ and $\phi_n^{\text{RSW2}} \geq \phi_n^{\text{CCK2}}$. Rows correspond to different values of $p_n \in \{40, 100, 200\}$ and $\rho \in \{0, 0.05, 0.9\}$. In all designs, we use 10,000 replications and 1,000 bootstrap samples. We emphasize that we employ the same bootstrap samples for all tests. We also note that there are no appreciable differences in the computation time of each test – e.g., computing one hundred replications of Model 1 with $\rho = 0$, $p = 200$, and one thousand bootstrap draws took 97.303 seconds for ϕ_n^{RSW} and 95.202 seconds for ϕ_n^{CCK} on a single Intel Core i5-8500 3.00GHz CPU.

We summarize our findings from the simulations as follows:

- Both ϕ_n^{RSW} and ϕ_n^{CCK} exhibit good size control even in settings where p_n exceeds the sample size $n = 100$, but ϕ_n^{CCK} tends to under-reject the null hypothesis more severely than ϕ_n^{RSW}. See, e.g., Model 2, $p = 200$, $\rho = 0$, and $\theta = 0$, in which case ϕ_n^{CCK} has rejection probability 0.60%, whereas ϕ_n^{RSW} has rejection probability 4.54%. In contrast, the tests ϕ_n^{RSW2} and ϕ_n^{CCK2} have considerably worse size control, over-rejecting the null hypothesis in some cases quite severely. See, e.g., Model 3, $p_n = 200$, $\rho = 0$, and $\theta = 0$, in which case ϕ_n^{RSW2} has rejection probability 6.79% and ϕ_n^{CCK2} has rejection probability 7.31%.

- The tests ϕ_n^{RSW2} and ϕ_n^{CCK2} are generally more powerful than ϕ_n^{RSW} and ϕ_n^{CCK}, but this feature must be weighed against their considerably worse size control. The test ϕ_n^{RSW} is generally at least as powerful as ϕ_n^{CCK}, and, at times, quite a bit more powerful. These instances tend to coincide with the values of p_n and ρ for which ϕ_n^{CCK} under-rejects the null hypothesis. See, e.g., Model 2, $p_n = 200$, $\rho = 0$, and $\theta = 0.2$, in which case ϕ_n^{CCK} has rejection probability only 26.69%, whereas ϕ_n^{RSW} has rejection probability 66.77%. The comparison between ϕ_n^{CCK2} and ϕ_n^{RSW2} is qualitatively similar.

- In nearly every replication, ϕ_n^{RSW} rejects the null hypothesis whenever ϕ_n^{CCK} does and ϕ_n^{RSW2} rejects the null hypothesis whenever ϕ_n^{CCK2} does. These results suggest that even though the analysis in Allen (2018) require the use of a Gaussian multiplier bootstrap, they may also hold approximately when employing the empirical bootstrap.

We conclude our simulation study by examining the extent to which the comparisons described above are artifacts of the differing choices of β used in implementing the various tests. To this end, we computed for each specification the rejection probabilities of all four tests at each $\beta \in [0.001, 0.025]$ in increments of 0.001. The results differ qualitatively depending on whether the specification corresponds to Models 1 and 3 (in which there are no components of $E_P[X_i]$ that are strictly negative both under the null and alternative)

Results for Model 1

		$\theta = 0$						$\theta = 0.2$					
		RSW	**RSW2**	**CCK**	**CCK2**	**$\geq$CCK**	**$\geq$CCK2**	**RSW**	**RSW2**	**CCK**	**CCK2**	**$\geq$CCK**	**$\geq$CCK2**
$p = 40$	$\rho = 0$	4.26	6.07	4.58	6.48	99.68	99.59	19.03	23.31	19.80	24.28	99.23	99.03
	$\rho = 0.5$	4.62	5.96	4.91	6.24	99.71	99.72	12.52	15.01	13.02	15.54	99.50	99.47
	$\rho = 0.9$	4.60	5.82	4.84	6.13	99.76	99.69	10.95	12.97	11.50	13.53	99.45	99.44
$p = 100$	$\rho = 0$	4.51	6.92	4.84	7.31	99.67	99.61	24.04	30.81	25.06	31.94	98.98	98.87
	$\rho = 0.5$	4.24	6.30	4.61	6.63	99.63	99.67	13.86	17.79	14.51	18.61	99.35	99.18
	$\rho = 0.9$	4.64	6.31	4.96	6.69	99.68	99.62	11.69	14.92	12.25	15.62	99.44	99.30
$p = 200$	$\rho = 0$	4.20	7.12	4.41	7.59	99.79	99.53	28.33	38.11	29.45	39.40	98.88	98.71
	$\rho = 0.5$	4.37	6.59	4.64	7.02	99.73	99.57	15.04	20.11	15.90	20.94	99.14	99.17
	$\rho = 0.9$	4.57	6.69	4.94	6.97	99.63	99.72	13.25	17.04	13.75	17.68	99.50	99.36

Results for Model 2

		$\theta = 0$						$\theta = 0.2$					
		RSW	**RSW2**	**CCK**	**CCK2**	**$\geq$CCK**	**$\geq$CCK2**	**RSW**	**RSW2**	**CCK**	**CCK2**	**$\geq$CCK**	**$\geq$CCK2**
$p = 40$	$\rho = 0$	4.34	4.87	0.65	1.16	100.00	100.00	45.94	48.72	16.82	20.68	100.00	100.00
	$\rho = 0.5$	3.62	4.30	0.73	0.89	100.00	100.00	22.33	26.22	9.69	11.92	100.00	100.00
	$\rho = 0.9$	2.19	3.01	0.67	0.87	100.00	100.00	13.60	16.82	8.23	9.87	100.00	99.99
$p = 100$	$\rho = 0$	4.50	5.56	0.69	1.17	100.00	100.00	58.33	63.95	22.22	28.39	100.00	100.00
	$\rho = 0.5$	3.11	4.20	0.78	1.02	100.00	100.00	24.36	30.86	12.03	15.48	100.00	100.00
	$\rho = 0.9$	1.79	2.79	0.70	1.07	100.00	100.00	14.36	19.35	10.06	12.63	99.97	100.00
$p = 200$	$\rho = 0$	4.54	5.80	0.60	1.17	100.00	100.00	66.77	74.45	26.69	35.41	100.00	100.00
	$\rho = 0.5$	3.05	4.50	0.69	1.14	100.00	100.00	26.03	34.57	14.43	19.44	100.00	100.00
	$\rho = 0.9$	1.57	2.81	0.77	1.22	100.00	100.00	15.60	21.94	11.90	15.74	100.00	100.00

Results for Model 3

		$\theta = 0$						$\theta = 0.2$					
		RSW	**RSW2**	**CCK**	**CCK2**	**$\geq$CCK**	**$\geq$CCK2**	**RSW**	**RSW2**	**CCK**	**CCK2**	**$\geq$CCK**	**$\geq$CCK2**
$p = 40$	$\rho = 0$	4.43	6.25	4.80	6.64	99.63	99.61	18.80	22.92	19.64	23.83	99.16	99.09
	$\rho = 0.5$	4.52	5.88	4.78	6.16	99.74	99.72	18.14	21.55	18.80	22.19	99.34	99.36
	$\rho = 0.9$	4.89	6.06	5.22	6.39	99.67	99.67	19.16	21.65	19.91	22.57	99.25	99.08
$p = 100$	$\rho = 0$	4.48	7.10	4.76	7.51	99.72	99.59	23.43	30.54	24.41	31.58	99.02	98.96
	$\rho = 0.5$	4.52	6.83	4.83	7.15	99.69	99.68	21.38	27.08	22.20	28.03	99.18	99.05
	$\rho = 0.9$	4.29	6.15	4.61	6.54	99.68	99.61	18.51	22.55	19.18	23.43	99.33	99.12
$p = 200$	$\rho = 0$	4.04	6.79	4.27	7.31	99.77	99.48	28.54	38.48	29.69	39.58	98.85	98.90
	$\rho = 0.5$	4.46	7.08	4.76	7.53	99.70	99.55	25.51	33.66	26.31	34.84	99.20	98.82
	$\rho = 0.9$	4.45	6.81	4.80	7.16	99.65	99.65	19.21	24.85	20.10	25.76	99.11	99.09

Results for Model 4

		$\theta = 0$						$\theta = 0.2$					
		RSW	**RSW2**	**CCK**	**CCK2**	**$\geq$CCK**	**$\geq$CCK2**	**RSW**	**RSW2**	**CCK**	**CCK2**	**$\geq$CCK**	**$\geq$CCK2**
$p = 40$	$\rho = 0$	4.70	5.43	0.89	1.41	100.00	100.00	45.39	48.69	16.87	20.69	100.00	100.00
	$\rho = 0.5$	4.46	4.84	0.78	1.48	100.00	100.00	40.66	43.51	15.02	17.96	100.00	100.00
	$\rho = 0.9$	4.72	5.15	1.15	1.57	100.00	100.00	41.54	43.65	17.43	20.55	100.00	100.00
$p = 100$	$\rho = 0$	4.62	5.57	0.65	1.15	100.00	100.00	58.12	63.70	21.91	28.03	100.00	100.00
	$\rho = 0.5$	4.40	5.24	0.62	1.16	100.00	100.00	48.96	53.78	19.28	24.01	100.00	100.00
	$\rho = 0.9$	4.26	4.88	0.73	1.18	100.00	100.00	40.77	44.06	16.81	20.05	100.00	100.00
$p = 200$	$\rho = 0$	4.27	5.71	0.67	1.17	100.00	100.00	66.11	73.39	26.98	35.80	100.00	100.00
	$\rho = 0.5$	4.16	5.37	0.61	0.99	100.00	100.00	55.93	62.82	22.42	29.96	100.00	100.00
	$\rho = 0.9$	4.57	5.53	0.67	1.09	100.00	100.00	41.44	45.90	16.23	20.57	100.00	100.00

Table 1: Rejection probabilities and percentage of replications for which $\phi_n^{\text{RSW}} \geq \phi_n^{\text{CCK}}$ and $\phi_n^{\text{RSW2}} \geq \phi_n^{\text{CCK2}}$. Tests ϕ_n^{RSW} and ϕ_n^{RSW2} implemented with $\beta = 0.005$ and ϕ_n^{CCK} and ϕ_n^{CCK2} implemented with $\beta = 0.001$.

or Models 2 and 4 (in which many components of $E_P[X_i]$ are strictly negative both under the null and alternative). We therefore only display one specification for each of these two sets of results.

Figures 1–2 display the results for Models 1 and 2 with $p = 100$ and $\rho = 0$. In each figure, the panel on the left corresponds to the case where $\theta = 0$ and the panel on the right corresponds to the case where

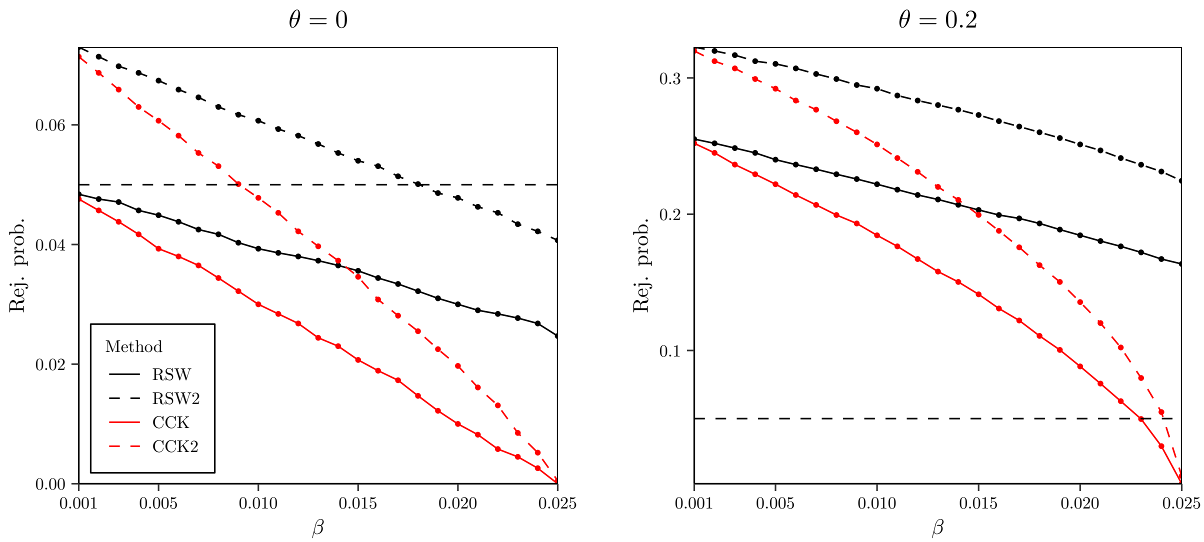

Figure 1: Rej. prob. for $\beta \in \{0.001, \ldots, 0.025\}$ in Model 1, $p = 100$, $\rho = 0$.

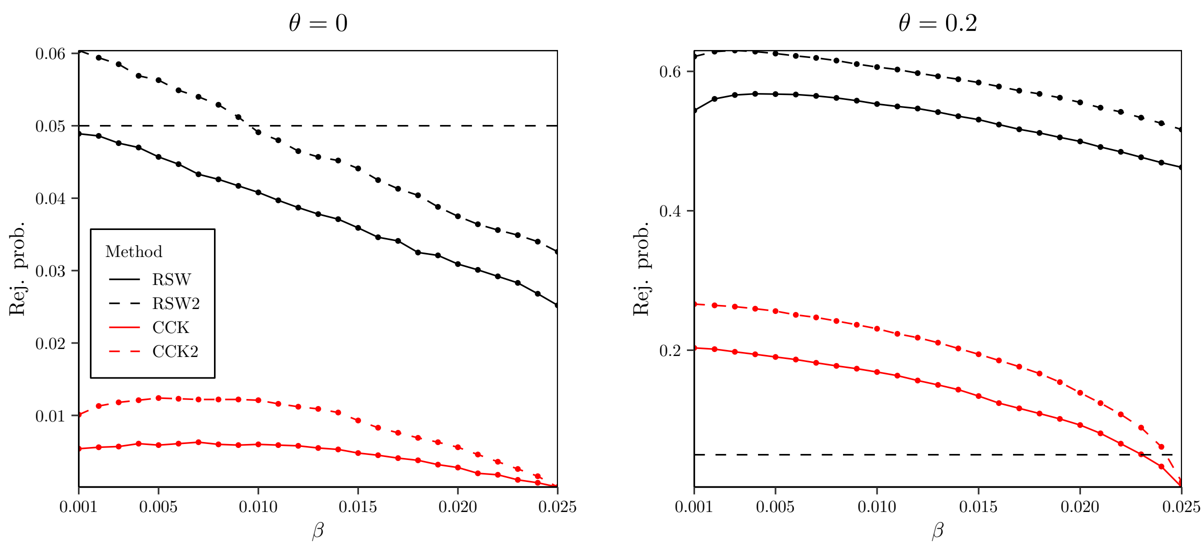

Figure 2: Rej. prob. for $\beta \in \{0.001, \ldots, 0.025\}$ in Model 2, $p = 100$, $\rho = 0$.

$\theta = 0.2$. Both figures provide further evidence that for any common choice of β, ϕ_n^{RSW} rejects more often than ϕ_n^{CCK} and ϕ_n^{RSW2} rejects more often than ϕ_n^{CCK2}. For all tests, in Model 1, the choice of β that leads to the most powerful test is given by the smallest choice of β – intuitively, in Model 1 there are no components of $E_P[X_i]$ that are strictly negative and hence implementing moment selection (setting $\beta > 0$) does not lead to a power gain. For the specification under Model 1, we therefore see that for any given choice of β for ϕ_n^{RSW}, there is a smaller choice of β for ϕ_n^{CCK} under which the two tests have similar rejection probabilities under both the null and alternative hypothesis. The same is true for ϕ_n^{RSW2} and ϕ_n^{CCK2}. In accord with our discussion in Section 2.1, however, we do see the power differences between the tests increase with β.

The results differ sharply for the specification under Model 2 (displayed in Figure 2). In that case, for any choice of β for ϕ_n^{RSW}, there is no choice of β for ϕ_n^{CCK} that makes ϕ_n^{CCK} more powerful: Indeed, the maximum rejection probability of ϕ_n^{CCK} over all values of β considered is smaller than the minimum rejection probability of ϕ_n^{RSW} across all values of β considered. The same is true for ϕ_n^{RSW2} and ϕ_n^{CCK2}. In accord with our discussion in Section 2.1, these power differences manifest themselves in a setting in which the alternative value for $E_P[X_i]$ has multiple strictly negative components.

4 Appendix

Proof of Theorem 2.1: For any vector $(\lambda_1, \ldots, \lambda_{p_n})' \equiv \lambda \in \mathbf{R}^{p_n}$, measure P, and $x \in \mathbf{R}$ define

$$\begin{aligned} F_n(x, \lambda, P) &\equiv P\left\{0 \vee \sqrt{n}(\bar{X}_{j,n} - \mu_j(P) + \lambda_j) \leq x S_{j,n} \text{ for all } 1 \leq j \leq p_n\right\} \\ J_n(x, \lambda, P) &\equiv P\left\{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P) + \lambda_j) \leq x S_{j,n} \text{ for all } 1 \leq j \leq p_n\right\}, \end{aligned}$$

and for any function $f : \mathbf{R} \to [0,1]$ let $f^{-1}(x) \equiv \inf\{c : f(c) \geq x\}$ with $f^{-1}(x) = +\infty$ whenever $\{c : f(c) \geq x\}$ is empty. Further define the event $\Omega_n(P)$ according to

$$\Omega_n(P) \equiv \{\mu_j(P) \leq \hat{u}_{j,n} \text{ for all } 1 \leq j \leq p_n\}, \tag{15}$$

and note that, for $(\hat{u}_{1,n}, \ldots, \hat{u}_{p_n,n})' \equiv \hat{u}_n \in \mathbf{R}^{p_n}$, the event $\Omega_n(P)$ implies $F_n(x, \mu(P), \hat{P}_n) \geq F_n(x, \hat{u}_n, \hat{P}_n)$ for all $x \in \mathbf{R}$, which yields $F_n^{-1}(x, \mu(P), \hat{P}_n) \leq F_n^{-1}(x, \hat{u}_n, \hat{P}_n)$ for all $x \in [0,1]$. In particular, by definition of $\hat{c}_n^{(2)}(1-\alpha+\beta)$ we obtain that $\Omega_n(P)$ implies $F_n^{-1}(1-\alpha+\beta, \mu(P), \hat{P}_n) \leq \hat{c}_n^{(2)}(1-\alpha+\beta)$, and hence Lemma 4.1 yields

$$\begin{aligned} \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_{0,n}} P\left\{T_n > \hat{c}_n^{(2)}(1-\alpha+\beta)\right\} &\leq \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_{0,n}} P\left\{T_n > \hat{c}_n^{(2)}(1-\alpha+\beta);\ \Omega_n(P)\right\} + \beta \\ &\leq \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_{0,n}} P\left\{T_n > F_n^{-1}(1-\alpha+\beta, \mu(P), \hat{P}_n)\right\} + \beta. \end{aligned} \tag{16}$$

Next, note that $S_{j,n} \geq 0$ almost surely implies $F_n(x, \lambda, P) = J_n(x, \lambda, P)$ for any λ, P, and $x \geq 0$, while for any λ, P and $x < 0$ we have $F_n(x, \lambda, P) \leq P\{S_{j,n} = 0 \text{ for all } 1 \leq j \leq p_n\}$. Hence, it follows that

$$\begin{aligned} &\sup_{x \in \mathbf{R}} \left|F_n(x, \mu(P), P) - F_n(x, \mu(P), \hat{P}_n)\right| \\ &\qquad \leq \sup_{x \geq 0} \left|J_n(x, \mu(P), P) - J_n(x, \mu(P), \hat{P}_n)\right| + P\left\{\max_{1 \leq j \leq p_n} S_{j,n} = 0\right\} + \hat{P}_n\left\{\max_{1 \leq j \leq p_n} S_{j,n} = 0\right\}, \end{aligned}$$

which together with Lemmas 4.2 and 4.3 implies there are sequence $\xi_n \downarrow 0$ and $\delta_n \downarrow 0$ such that

$$\inf_{P \in \mathbf{P}_n} P\left\{\sup_{x \in \mathbf{R}} \left|F_n(x, \mu(P), P) - F_n(x, \mu(P), \hat{P}_n)\right| \leq \xi_n\right\} \geq 1 - \delta_n. \tag{17}$$

Moreover, since $F_n(F_n^{-1}(1-\alpha+\beta, \mu(P), \hat{P}_n), \mu(P), \hat{P}_n) \geq 1-\alpha+\beta$, it follows that

$$\begin{aligned} \left\{\sup_{x \in \mathbf{R}} \left|F_n(x, \mu(P), P) - F_n(x, \mu(P), \hat{P}_n)\right| \leq \xi_n\right\} &\subseteq \left\{F_n(F_n^{-1}(1-\alpha+\beta, \mu(P), \hat{P}_n), \mu(P), P) \geq 1-\alpha+\beta-\xi_n\right\} \\ &\subseteq \left\{F_n^{-1}(1-\alpha+\beta, \mu(P), \hat{P}_n) \geq F_n^{-1}(1-\alpha+\beta-\xi_n, \mu(P), P)\right\}. \end{aligned} \tag{18}$$

Thus, since $P\{T_n \leq x\} = F_n(x, \mu(P), P)$, results (17) and (18) together establish that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_{0,n}} P\left\{T_n > F_n^{-1}(1-\alpha+\beta,\mu(P),\hat{P}_n)\right\}$$
$$\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_{0,n}} P\left\{T_n > F_n^{-1}(1-\alpha+\beta-\xi_n,\mu(P),P)\right\} + \delta_n \leq \limsup_{n\to\infty} \alpha-\beta-\xi_n+\delta_n. \quad (19)$$

The claim of the theorem therefore follows from (16), (19), $\xi_n \downarrow 0$, and $\delta_n \downarrow 0$. ■

Lemma 4.1. *Let Assumption 2.1 hold. If $\beta \in (0, 0.5)$, then it follows that*

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_{0,n}} P\left\{\mu_j(P) \leq \hat{u}_{j,n} \text{ for all } 1 \leq j \leq p_n\right\} \geq 1-\beta.$$

Proof: The proof follows from Lemma 4.2 and arguments in the proof of Lemma A.1 in Romano and Shaikh (2012). First note that for any $P \in \mathbf{P}_{0,n}$ we have $\mu_j(P) \leq 0$ for all $1 \leq j \leq p_n$, and therefore by definition of $\hat{u}_{j,n}$

$$P\left\{\mu_j(P) \leq \hat{u}_{j,n} \text{ for all } 1 \leq j \leq p_n\right\} = P\left\{\sqrt{n}(\mu_j(P)-\bar{X}_{j,n}) \leq S_{j,n}\hat{c}_n^{(1)}(1-\beta) \text{ for all } 1 \leq j \leq p_n\right\}. \quad (20)$$

Next, for any measure P we define the function $F_n(\cdot,P) : \mathbf{R} \to [0,1]$ to be given by

$$F_n(x,P) \equiv P\left\{\sqrt{n}(\mu_j(P)-\bar{X}_{j,n}) \leq S_{j,n}x \text{ for all } 1 \leq j \leq p_n\right\}. \quad (21)$$

Then note that if $\{X_i\}_{i=1}^n$ satisfies Assumption 2.1, then so does $\{-X_i\}_{i=1}^n$. Hence, we may apply Lemma 4.2 to conclude there exist sequences $\xi_n \downarrow 0$ and $\delta_n \downarrow 0$ such that

$$\inf_{P\in\mathbf{P}_n} P\left\{\sup_{x\geq 0}\left|F_n(x,P)-F_n(x,\hat{P}_n)\right| \leq \xi_n\right\} \geq 1-\delta_n. \quad (22)$$

Further let Φ denote the c.d.f. of a standard normal random variable and note that Theorem 1.1. Bentkus and Götze (1996) and Assumption 2.1(iii) imply

$$\sup_{P\in\mathbf{P}_n} F_n(0,P) \leq \sup_{P\in\mathbf{P}_n} P\left\{\sqrt{n}(\mu_1(P)-\bar{X}_{1,n}) \leq S_{1,n}\times 0\right\} \leq 0.5 + \frac{KM_{1,n}}{\sqrt{n}} \quad (23)$$

for some finite constant $K \in \mathbf{R}$. Next, for any $f : \mathbf{R} \to [0,1]$ let $f^{-1}(x) \equiv \inf\{c : f(c) \geq x\}$ with $f^{-1}(x) = +\infty$ if $\{c : f(c) \geq x\} = \emptyset$, and define the event $\Omega_n(P)$ to be given by

$$\Omega_n(P) \equiv \left\{\sup_{x\geq 0}\left|F_n(x,P)-F_n(x,\hat{P}_n)\right| \leq \xi_n\right\}. \quad (24)$$

Then note that since $\beta < 0.5$ and $M_{1,n}/\sqrt{n} = o(1)$ by hypothesis, result (23) implies that

$$\sup_{P\in\mathbf{P}_n} F_n(0,P) + \xi_n < 1-\beta \quad (25)$$

for n sufficiently large. Therefore, the definitions of $\hat{c}_n^{(1)}(1-\beta)$ and $\Omega_n(P)$ yield

$$\Omega_n(P) \subseteq \{F_n(0,\hat{P}_n) < 1-\beta\} \subseteq \{\hat{c}_n^{(1)}(1-\beta) \geq 0\} \quad (26)$$

for n sufficiently large. Combining definition (24) and result (26) further implies

$$\Omega_n(P) \subseteq \left\{F_n(\hat{c}_n^{(1)}(1-\beta),P) \geq F_n(\hat{c}_n^{(1)}(1-\beta),\hat{P}_n)-\xi_n\right\}$$
$$\subseteq \left\{F_n(\hat{c}_n^{(1)}(1-\beta),P) \geq 1-\beta-\xi_n\right\} \subseteq \left\{\hat{c}_n^{(1)}(1-\beta) \geq F_n^{-1}(1-\beta-\xi_n,P)\right\}, \quad (27)$$

where the second and third set inclusions follow by definition of $\hat{c}_n^{(1)}(1-\beta)$ and $F_n^{-1}(\cdot, P)$. Hence, results (20), (22), and the definitions of $F_n^{-1}(\cdot, P)$ and $\Omega_n(P)$ yield

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_{0,n}} P\left\{\mu_j(P) \le \hat{u}_{j,n}\ \forall 1\le j\le p_n\right\} \\
&\qquad \ge \liminf_{n\to\infty} \inf_{P\in\mathbf{P}_n} P\left\{\sqrt{n}(\mu_j(P) - \bar{X}_{j,n}) \le S_{j,n} F_n^{-1}(1-\beta-\xi_n, P)\ \forall 1\le j\le p_n\right\} - \delta_n \\
&\qquad \ge \liminf_{n\to\infty} 1-\beta-\xi_n-\delta_n,
\end{aligned}$$

which establishes the claim of the lemma because $\xi_n \downarrow 0$ and $\delta_n \downarrow 0$. ■

Lemma 4.2. *Let Assumption 2.1 hold and for any $(\lambda_1, \ldots, \lambda_{p_n})' \equiv \lambda \in \mathbf{R}_-^{p_n}$, $P \in \mathbf{P}_n$, and $x \in \mathbf{R}$ define*

$$J_n(x, \lambda, P) \equiv P\left\{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P) + \lambda_j) \le x S_{j,n} \text{ for all } 1\le j\le p_n\right\}.$$

Then, there exists a sequence $\xi_n \downarrow 0$ such that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_n} P\left\{\sup_{x\ge 0}\sup_{\lambda\in\mathbf{R}_-^{p_n}} \left|J_n(x,\lambda,\hat{P}_n) - J_n(x,\lambda,P)\right| \le \xi_n\right\} = 1.$$

Proof: We first note that $\sigma_j(P) > 0$ for all $1\le j\le p_n$ by Assumption 2.1(ii) implies that

$$\begin{aligned}
J_n(x,\lambda,P) &= P\left\{\frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P))}{\sigma_j(P)} \le x\frac{S_{j,n}}{\sigma_j(P)} - \frac{\sqrt{n}\lambda_j}{\sigma_j(P)} \text{ for all } 1\le j\le p_n\right\} \\
J_n(x,\lambda,\hat{P}_n) &= \hat{P}_n\left\{\frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(\hat{P}_n))}{\sigma_j(P)} \le x\frac{S_{j,n}}{\sigma_j(P)} - \frac{\sqrt{n}\lambda_j}{\sigma_j(P)} \text{ for all } 1\le j\le p_n\right\}.
\end{aligned}$$

Next, let $(Z_1,\ldots,Z_{p_n})' \equiv Z \in \mathbf{R}^{p_n}$ be a Gaussian vector satisfying $E[Z_j] = 0$ and $E[Z_jZ_k] = E_P[(X_{i,j}-\mu_j(P))(X_{i,k}-\mu_k(P))]/\sigma_j(P)\sigma_k(P)$ for any $1\le j,k\le p_n$, and for any measure P, $(\lambda_1,\ldots,\lambda_{p_n})' \equiv \lambda \in \mathbf{R}_-^{p_n}$ and $(\omega_1,\ldots,\omega_{p_n})' \equiv \omega \in \mathbf{R}^{p_n}$ satisfying $\omega_j > 0$ for all $1\le j\le p_n$, define $F_n(x,\lambda,\omega,P)$ and $G_n(x,\lambda,\omega,P)$ to equal

$$F_n(x,\lambda,\omega,P) \equiv P\left\{\frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P))}{\omega_j} \le x - \frac{\sqrt{n}\lambda_j}{\omega_j} \text{ for all } 1\le j\le p_n\right\} \tag{28}$$

$$G_n(x,\lambda,\omega,P) \equiv P\left\{Z_j \le x - \frac{\sqrt{n}\lambda_j}{\omega_j} \text{ for all } 1\le j\le p_n\right\}. \tag{29}$$

Since $B_n^2\log^{3.5}(p_n)/n^{(1-\delta)/2} = o(1)$ for some $\delta > 0$ by Assumption 2.1(v), we may find an $\epsilon_n \downarrow 0$ satisfying

$$\frac{B_n^2\log^2(p_n)}{n^{(1-\delta)/2}} = o(\epsilon_n) \qquad\qquad \log(p_n)\epsilon_n = o(1).$$

In particular, the condition $B_n^2\log^2(p_n)/n^{(1-\delta)/2} = o(\epsilon_n)$ implies that the sequence η_n defined by

$$\eta_n \equiv \sup_{P\in\mathbf{P}_n} P\left\{\max_{1\le j\le p_n}\left|\frac{S_{j,n}}{\sigma_j(P)} - 1\right| > \epsilon_n\right\} \tag{30}$$

satisfies $\eta_n = o(1)$ by Lemma 4.3(i). Moreover, by definitions (28) and (30) we can conclude that

$$F_n(x(1-\epsilon_n),\lambda,\sigma(P),P) - \eta_n \le J_n(x,\lambda,P) \le F_n(x(1+\epsilon_n),\lambda,\sigma(P),P) + \eta_n \tag{31}$$

for all $x \ge 0$, $P\in\mathbf{P}_n$, and $\lambda\in\mathbf{R}_-^{p_n}$. Next note $(M_{1,n}^2 \vee M_{2,n}^2 \vee B_n^2)\log^{3.5}(p_n n)/\sqrt{n} = o(1)$ by Assumption 2.1(v), Assumptions 2.1(i)(iii)(iv) and Proposition 2.1 in Chernozhukov et al. (2017) imply that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} \sup_{x\in\mathbf{R}} \sup_{\lambda\in\mathbf{R}_-^{p_n}} |F_n(x,\lambda,\sigma(P),P) - G_n(x,\lambda,\sigma(P),P)| = 0. \tag{32}$$

On the other hand, we may further conclude by Lemma 4.4 and $\epsilon_n \log(p_n) = o(1)$ by construction that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} \sup_{x\geq 0} \sup_{\lambda\in\mathbf{R}_-^{p_n}} G_n((1+\epsilon_n)x,\lambda,\sigma(P),P) - G_n((1-\epsilon_n)x,\lambda,\sigma(P),P)$$
$$\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} \sup_{x\geq 0} \sup_{\lambda\in\mathbf{R}_-^{p_n}} P\left\{\left|\max_{1\leq j\leq p_n} Z_j + \frac{\sqrt{n}\lambda_j}{\sigma_j(P)} - x\right| \leq 2\epsilon_n x\right\} = 0. \tag{33}$$

Therefore, combining results (30), (31), (32), and (33) and employing that $\eta_n = o(1)$ we obtain

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} \sup_{x\geq 0} \sup_{\lambda\in\mathbf{R}_-^{p_n}} |J_n(x,\lambda,P) - G_n(x,\lambda,\sigma(P),P)| = 0. \tag{34}$$

To conclude the proof, we set $\bar{M}_n \equiv M_{1,n} \vee M_{2,n} \vee B_n$ and define the events $\Omega_{1,n}(P)$ and $\Omega_{2,n}(P)$ according to

$$\Omega_{1,n}(P) \equiv \left\{P\left\{\max_{1\leq j\leq p_n}\left|\frac{S^*_{j,n}}{\sigma_j(P)} - 1\right| > \epsilon_n \Big| \{X_i\}_{i=1}^n\right\} \leq \frac{K}{n^\delta}\right\}$$
$$\Omega_{2,n}(P) \equiv \left\{\sup_{x\in\mathbf{R}} \sup_{\lambda\in\mathbf{R}_-^{p_n}} \left|F_n(x,\lambda,\sigma(P),\hat{P}_n) - G_n(x,\lambda,\sigma(P),P)\right| \leq K\left(\frac{\bar{M}_n^2\log^{3.5}(p_n n)}{n^{(1-\delta)/2}}\right)^{1/6}\right\}$$

and note that for $\Omega_n(P) \equiv \Omega_{1,n}(P) \cap \Omega_{2,n}(P)$, for appropriately selected $K < \infty$, Lemma 4.3(ii) and Proposition 4.3 in Chernozhukov et al. (2017) (applied with $\alpha = n^{-\delta}$) allow us to conclude that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_n} P\{\Omega_n(P)\} = 1. \tag{35}$$

Furthermore, observe that under $\Omega_n(P)$ we may argue as in result (31) to obtain that for all $x \geq 0$ and $\lambda \in \mathbf{R}_-^{p_n}$

$$J_n(x,\lambda,\hat{P}_n) \leq F_n((1+\epsilon_n)x,\lambda,\sigma(P),\hat{P}_n) + \frac{K}{n^\delta}$$
$$J_n(x,\lambda,\hat{P}_n) \geq F_n((1-\epsilon_n)x,\lambda,\sigma(P),\hat{P}_n) - \frac{K}{n^\delta}.$$

Therefore, employing results (33) and (35) imply that there exists a sequence $\xi_n \downarrow 0$ such that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_n} P\left\{\sup_{x\geq 0} \sup_{\lambda\in\mathbf{R}_-^{p_n}} |J_n(x,\lambda,\hat{P}_n) - G_n(x,\lambda,\sigma(P),P)| \leq \xi_n\right\} = 1. \tag{36}$$

The lemma thus follows from results (34) and (36). ■

Lemma 4.3. *Let Assumption 2.1(i)(ii)(iv) hold. Then: (i) For any sequence $\epsilon_n \downarrow 0$ satisfying $B_n^2\log^2(p_n)/n^{(1-\delta)/2} = o(\epsilon_n)$ for some $\delta \in (0,1)$ it follows that*

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} P\left\{\max_{1\leq j\leq p_n}\left|\frac{S_{j,n}}{\sigma_j(P)} - 1\right| > \epsilon_n\right\} = 0. \tag{37}$$

(ii) For any $\epsilon_n \downarrow 0$ satisfying the condition of part (i) there is a $K < \infty$ such that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} P\left\{P\left\{\max_{1\leq j\leq p_n}\left|\frac{S^*_{j,n}}{\sigma_j(P)} - 1\right| > \epsilon_n \Big| \{X_i\}_{i=1}^n\right\} \leq \frac{K}{n^\delta}\right\} = 1.$$

Proof: The first claim of the lemma corresponds to Lemma D.5 in Chernozhukov et al. (2019), which we may apply by Assumptions 2.1(i)(ii)(iv). In order to establish the second claim of the lemma we first define the event

$$\Omega_{1,n}(P) \equiv \left\{ \max_{1\le j\le p_n} \left| \frac{S_{j,n}}{\sigma_j(P)} - 1 \right| \le \frac{\epsilon_n}{2} \right\},$$

where ϵ_n satisfies $B_n^2 \log^2(p_n)/n^{(1-\delta)/2} = o(\epsilon_n)$ for some $\delta \in (0,1)$ by hypothesis. We further define $\hat{B}_n^4 \in \mathbf{R}$ to equal

$$\hat{B}_n^4 \equiv \frac{1}{n}\sum_{i=1}^n \max_{1\le j\le p_n} \left(\frac{X_{i,j} - \bar{X}_{j,n}}{S_{j,n}} \right)^4$$

and note that since $\epsilon_n \downarrow 0$ it follows that, for n sufficiently large, $\Omega_{1,n}(P)$ implies $S_{j,n}$ is positive for all $1 \le j \le p_n$. Furthermore, Lemma D.5 in Chernozhukov et al. (2019) implies there are finite positive $K_1, K_2 \in \mathbf{R}$ satisfying

$$I\{\Omega_{1,n}(P)\} \times P\left\{ \max_{1\le j\le p_n} \left| \frac{S_{j,n}^*}{S_{j,n}} - 1 \right| > K_1 \frac{\hat{B}_n^2 \log^2(p_n)}{n^{(1-\delta)/2}} \Big| \{X_i\}_{i=1}^n \right\} \le I\{\Omega_{1,n}(P)\} \times \frac{K_2}{n^\delta}. \quad (38)$$

Moreover, the definition of the event $\Omega_{1,n}(P)$ and the inequality $(a+b)^4 \le 8(a^4+b^4)$ also yield that

$$\begin{aligned} I\{\Omega_{1,n}(P)\} \times \hat{B}_n^4 &\le I\{\Omega_{1,n}(P)\} \times \max_{1\le j\le p_n} \frac{\sigma_j^4(P)}{S_{j,n}^4} \times \frac{1}{n}\sum_{i=1}^n \max_{1\le j\le p_n} \left(\frac{X_{i,j} - \bar{X}_{j,n}}{\sigma_j(P)} \right)^4 \\ &\le 8\left(1 + \frac{\epsilon_n}{2}\right)^4 \times \frac{1}{n}\sum_{i=1}^n \left(\max_{1\le j\le p_n} \left(\frac{X_{i,j} - \mu_j(P)}{\sigma_j(P)} \right)^4 + \max_{1\le j\le p_n} \left(\frac{\bar{X}_{j,n} - \mu_j(P)}{\sigma_j(P)} \right)^4 \right). \quad (39) \end{aligned}$$

Next note that for any sequence $\ell_n \downarrow 0$, Assumption 2.1(iv) and Markov's inequality imply that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} P\left\{ \frac{1}{n}\sum_{i=1}^n \max_{1\le j\le p_n} \left(\frac{X_{i,j} - \mu_j(P)}{\sigma_j(P)} \right)^4 > \frac{B_n^4}{\ell_n} \right\} = 0. \quad (40)$$

Furthermore, since $B_n \ge 1$ by Jensen's inequality, we note that $\epsilon_n \downarrow 0$ and the condition $B_n^2 \log^2(p_n)/n^{(1-\delta)/2} = o(\epsilon_n)$ together imply that $\log^2(p_n)/n = o(1)$. Therefore, $\ell_n \downarrow 0$ and equation (73) in Chernozhukov et al. (2019) yield

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} P\left\{ \max_{1\le j\le p_n} \left| \frac{1}{n}\sum_{i=1}^n \frac{X_{i,j} - \mu_j(P)}{\sigma_j(P)} \right|^4 > \frac{B_n^4}{\ell_n} \right\} = 0. \quad (41)$$

Combining results (39), (40), (41), and that $P\{\Omega_{1,n}(P)\} = 1 + o(1)$ uniformly in $P \in \mathbf{P}_n$ by part (i) of this lemma, it follows that there exists a constant $K_3 < \infty$ independent of the sequence ℓ_n with

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} P\left\{ \hat{B}_n^4 > K_3 \frac{B_n^4}{\ell_n} \right\} = 0.$$

Thus, by selecting $\ell_n \downarrow 0$ to satisfy $B_n^2 \log^2(p_n)/(\sqrt{\ell_n} n^{(1-\delta)/2}) = o(\epsilon_n)$, which is possible due to $B_n^2 \log^2(p_n)/n^{(1-\delta)/2} = o(\epsilon_n)$ by hypothesis, we are able to conclude from result (38) that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} P\left\{ P\left\{ \max_{1\le j\le p_n} \left| \frac{S_{j,n}^*}{S_{j,n}} - 1 \right| > \frac{\epsilon_n}{4} \Big| \{X_i\}_{i=1}^n \right\} \le \frac{K_2}{n^\delta} \right\} = 1. \quad (42)$$

Finally, note that for any $(a_1, \ldots, a_{p_n})' \in \mathbf{R}^{p_n}$, we obtain by definition of the event $\Omega_{1,n}(P)$ that

$$I\{\Omega_{1,n}(P)\} \times \max_{1\le j\le p_n} \left| \frac{a_j}{\sigma_j(P)} - 1 \right| \le I\{\Omega_{1,n}(P)\} \times \left(\max_{1\le j\le p_n} \left| \frac{a_j}{S_{j,n}} - 1 \right| \frac{S_{j,n}}{\sigma_j(P)} + \max_{1\le j\le p_n} \left| \frac{S_{j,n}}{\sigma_j(P)} - 1 \right| \right)$$

$$\leq I\{\Omega_{1,n}(P)\} \times \left(\max_{1\leq j\leq p_n} \left| \frac{a_j}{S_{j,n}} - 1 \right| (1 + \frac{\epsilon_n}{2}) + \frac{\epsilon_n}{2} \right). \tag{43}$$

Thus, $P\{\Omega_{1,n}(P)\} = 1 + o(1)$ uniformly in $P \in \mathbf{P}_n$ by part (i) of this lemma, and results (42) and (43) imply

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} P\left\{ P\left\{ \max_{1\leq j\leq p_n} \left| \frac{S^*_{j,n}}{\sigma_j(P)} - 1 \right| > \epsilon_n \middle| \{X_i\}_{i=1}^n \right\} \leq \frac{K_2}{n^\delta} \right\}$$
$$\geq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} P\left\{ P\left\{ \max_{1\leq j\leq p_n} \left| \frac{S^*_{j,n}}{S_{j,n}} - 1 \right| > \frac{\epsilon_n}{4} \middle| \{X_i\}_{i=1}^n \right\} \leq \frac{K_2}{n^\delta} \right\} = 1,$$

which establishes the second claim of the lemma. ■

Lemma 4.4. *Let $(Z_1, \ldots, Z_p)' \equiv Z \in \mathbf{R}^p$ be Gaussian with $E[Z_j] = 0$ and $E[Z_j^2] = 1$ for all $1 \leq j \leq p$, and $(s_1, \ldots, s_p) \equiv s \in \mathbf{R}^p_-$. Then, there is a constant $C < \infty$ such that for all $\delta \in (0, 0.5]$ and $t > 0$:*

$$\sup_{x\geq 0} P\left\{ \left| \max_{1\leq j\leq p} (Z_j + s_j) - x \right| \leq \delta x \right\} \leq C\delta(1 + \sqrt{\log(p)} + t)^2 + \exp\left\{ -\frac{t^2}{2} \right\}.$$

Proof: Let m_p denote the median of $\max_{1\leq j\leq p} Z_j$, and note that by Kwapień (1994) $\mathrm{m}_p \leq E[\max_{1\leq j\leq p} Z_j]$. Since in addition $E[\max_{1\leq j\leq p} Z_j] \leq \sqrt{2\log(p)}$ by Lemmas 2.2.1 and 2.2.2 in van der Vaart and Wellner (1996), we obtain

$$\mathrm{m}_p \leq \sqrt{2\log(p)}. \tag{44}$$

Next, for any $t > 0$ we set $a \equiv 2(\sqrt{2\log(p)} + t)$ and observe the union bound allows us to conclude that

$$\sup_{0\leq x\leq a} P\left\{ \left| \max_{1\leq j\leq p} (Z_j + s_j) - x \right| \leq \delta x \right\} \leq \sup_{0\leq x\leq a} P\left\{ \left| \max_{1\leq j\leq p: s_j\leq -a/2} (Z_j + s_j) - x \right| \leq \delta x \right\}$$
$$+ \sup_{0\leq x\leq a} P\left\{ \left| \max_{1\leq j\leq p: s_j > -a/2} (Z_j + s_j) - x \right| \leq \delta x \right\}. \tag{45}$$

Moreover, we note that $\delta \in (0, 0.5]$ and $x > 0$ imply $x(1 - \delta) > 0$, and hence we obtain

$$\sup_{0\leq x\leq a} P\left\{ \left| \max_{1\leq j\leq p: s_j\leq -a/2} (Z_j + s_j) - x \right| \leq \delta x \right\} \leq P\left\{ \max_{1\leq j\leq p: s_j\leq -a/2} (Z_j + s_j) \geq 0 \right\}$$
$$\leq P\left\{ \max_{1\leq j\leq p} Z_j \geq \sqrt{2\log(p)} + t \right\} \leq \exp\left\{ -\frac{t^2}{2} \right\}, \tag{46}$$

where the second inequality holds by definition of a, while the final inequality follows from Borell's inequality (see, e.g., the Corollary in pg. 82 of Davydov et al. (1998)), result (44), and $1 - \Phi(t) \leq \exp\{-t^2/2\}$ for any $t > 0$ and Φ the c.d.f. of a standard normal random variable. Next note that Lemma A.1 in Chernozhukov et al. (2017) yields

$$\sup_{0\leq x\leq a} P\left\{ \left| \max_{1\leq j\leq p: s_j > -a/2} (Z_j + s_j) - x \right| \leq \delta x \right\} \lesssim \delta a \sqrt{\log(p)}. \tag{47}$$

Moreover, since $s_j \leq 0$ for all $1 \leq j \leq p$ and $\delta \leq 0.5$ we can additionally conclude that

$$\sup_{x\geq a} P\left\{ \left| \max_{1\leq j\leq p} (Z_j + s_j) - x \right| \leq \delta x \right\} \leq \sup_{x\geq a} P\left\{ \max_{1\leq j\leq p} Z_j \geq x(1 - \delta) \right\} \leq P\left\{ \max_{1\leq j\leq p} Z_j \geq \frac{a}{2} \right\} \leq \exp\left\{ -\frac{t^2}{2} \right\}, \tag{48}$$

where the final inequality follows by another application of Borell's inequality and the arguments employed in (46). The lemma follows from (45), (46), (47), and (48). ■

References

Allen, R. (2018). Testing moment inequalities: Selection versus recentering. *Economics Letters*, **162** 124–126.

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

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

Bentkus, V. and Götze, F. (1996). The Berry-Esseen bound for Student's statistic. *The Annals of Probability*, **24** 491–503.

Canay, I. A. and Shaikh, A. M. (2017). Practical and Theoretical Advances in Inference for Partially Identified Models. In *Advances in Economics and Econometrics: Eleventh World Congress* (B. Honoré, A. Pakes, M. Piazzesi and L. Samuelson, eds.), vol. 2 of *Econometric Society Monographs*. Cambridge University Press, 271–306.

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

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, **45** 2309–2352.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *Review of Economic Studies (forthcoming)*.

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

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998). *Local properties of distributions of stochastic functionals*. American Mathematical Society.

Kwapień, S. (1994). A remark on the median and the expectation of convex functions of Gaussian vectors. In *Probability in Banach spaces, 9*. Springer, 271–272.

Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, **40** 2798–2822.

Romano, J. P., Shaikh, A. M. and Wolf, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica*, **82** 1979–2002.

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, **100** 94–108.

Romano, J. P. and Wolf, M. (2018). Multiple testing of one-sided hypotheses: combining Bonferroni and the bootstrap. In *International Conference of the Thailand Econometrics Society*. Springer, 78–94.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.