The Annals of Statistics
2012, Vol. 40, No. 6, 2798–2822
DOI: 10.1214/12-AOS1051

ON THE UNIFORM ASYMPTOTIC VALIDITY OF SUBSAMPLING AND THE BOOTSTRAP

BY JOSEPH P. ROMANO AND AZEEM M. SHAIKH

Stanford University and University of Chicago

This paper provides conditions under which subsampling and the bootstrap can be used to construct estimators of the quantiles of the distribution of a root that behave well uniformly over a large class of distributions **P**. These results are then applied (i) to construct confidence regions that behave well uniformly over **P** in the sense that the coverage probability tends to at least the nominal level uniformly over **P** and (ii) to construct tests that behave well uniformly over **P** in the sense that the size tends to no greater than the nominal level uniformly over **P**. Without these stronger notions of convergence, the asymptotic approximations to the coverage probability or size may be poor, even in very large samples. Specific applications include the multivariate mean, testing moment inequalities, multiple testing, the empirical process and U-statistics.

1. Introduction. Let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$, and denote by $J_n(x, P)$ the distribution of a real-valued root $R_n = R_n(X^{(n)}, P)$ under P. In statistics and econometrics, it is often of interest to estimate certain quantiles of $J_n(x, P)$. Two commonly used methods for this purpose are subsampling and the bootstrap. This paper provides conditions under which these estimators behave well uniformly over **P**. More precisely, we provide conditions under which subsampling and the bootstrap may be used to construct estimators $\hat{c}_n(\alpha_1)$ of the α_1 quantiles of $J_n(x, P)$ and $\hat{c}_n(1 - \alpha_2)$ of the $1 - \alpha_2$ quantiles of $J_n(x, P)$, satisfying

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}} P\{\hat{c}_n(\alpha_1) \le R_n \le \hat{c}_n(1 - \alpha_2)\} \ge 1 - \alpha_1 - \alpha_2. \tag{1}$$

Here, $\hat{c}_n(0)$ is understood to be $-\infty$, and $\hat{c}_n(1)$ is understood to be $+\infty$. For the construction of two-sided confidence intervals of nominal level $1 - 2\alpha$ for a real-valued parameter, we typically would consider $\alpha_1 = \alpha_2 = \alpha$, while for a one-sided confidence interval of nominal level $1 - \alpha$ we would consider either $\alpha_1 = 0$ and $\alpha_2 = \alpha$, or $\alpha_1 = \alpha$ and $\alpha_2 = 0$. In many cases, it is possible to replace the $\liminf_{n \to \infty}$ and $\ge$ in (1) with $\lim_{n \to \infty}$ and $=$, respectively. These results differ from those usually stated in the literature in that they require the convergence to hold uniformly over **P** instead of just pointwise over **P**. The importance of this

Received April 2012; revised September 2012.
MSC2010 subject classifications. 62G09, 62G10.
Key words and phrases. Bootstrap, empirical process, moment inequalities, multiple testing, subsampling, uniformity, U-statistic.

stronger notion of convergence when applying these results is discussed further below.

As we will see, the result (1) may hold with $\alpha_1 = 0$ and $\alpha_2 = \alpha \in (0, 1)$, but it may fail if $\alpha_2 = 0$ and $\alpha_1 = \alpha \in (0, 1)$, or the other way round. This phenomenon arises when it is not possible to estimate $J_n(x, P)$ uniformly well with respect to a suitable metric, but, in a sense to be made precise by our results, it is possible to estimate it sufficiently well to ensure that (1) still holds for certain choices of α_1 and α_2. Note that metrics compatible with the weak topology are not sufficient for our purposes. In particular, closeness of distributions with respect to such a metric does not ensure closeness of quantiles. See Remark 2.7 for further discussion of this point. In fact, closeness of distributions with respect to even stronger metrics, such as the Kolmogorov metric, does not ensure closeness of quantiles either. For this reason, our results rely heavily on Lemma A.1 which relates closeness of distributions with respect to a suitable metric and coverage statements.

In contrast, the usual arguments for the pointwise asymptotic validity of subsampling and the bootstrap rely on showing for each $P \in \mathbf{P}$ that $\hat{c}_n(1-\alpha)$ tends in probability under P to the $1-\alpha$ quantile of the limiting distribution of R_n under P. Because our results are uniform in $P \in \mathbf{P}$, we must consider the behavior of R_n and $\hat{c}_n(1-\alpha)$ under arbitrary sequences $\{P_n \in \mathbf{P} : n \geq 1\}$, under which the quantile estimators need not even settle down. Thus, the results are not trivial extensions of the usual pointwise asymptotic arguments.

The construction of $\hat{c}_n(\alpha)$ satisfying (1) is useful for constructing confidence regions that behave well uniformly over $\mathbf{P}$. More precisely, our results provide conditions under which subsampling and the bootstrap can be used to construct confidence regions $C_n = C_n(X^{(n)})$ of level $1-\alpha$ for a parameter $\theta(P)$ that are uniformly consistent in level in the sense that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P\{\theta(P) \in C_n\} \geq 1-\alpha. \tag{2}$$

Our results are also useful for constructing tests $\phi_n = \phi_n(X^{(n)})$ of level α for a null hypothesis $P \in \mathbf{P}_0 \subseteq \mathbf{P}$ against the alternative $P \in \mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0$ that are uniformly consistent in level in the sense that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} E_P[\phi_n] \leq \alpha. \tag{3}$$

In some cases, it is possible to replace the $\liminf_{n\to\infty}$ and $\geq$ in (2) or the $\limsup_{n\to\infty}$ and $\leq$ in (3) with $\lim_{n\to\infty}$ and $=$, respectively.

Confidence regions satisfying (2) are desirable because they ensure that for every $\varepsilon > 0$ there is an N such that for $n > N$ we have that $P\{\theta(P) \in C_n\}$ is no less than $1-\alpha-\varepsilon$ for all $P \in \mathbf{P}$. In contrast, confidence regions that are only pointwise consistent in level in the sense that

$$\liminf_{n\to\infty} P\{\theta(P) \in C_n\} \geq 1-\alpha$$

for each fixed $P \in \mathbf{P}$ have the feature that there exists some $\varepsilon > 0$ and $\{P_n \in \mathbf{P} : n \geq 1\}$ such that $P_n\{\theta(P_n) \in C_n\}$ is less than $1 - \alpha - \varepsilon$ infinitely often. Likewise, tests satisfying (3) are desirable for analogous reasons. For this reason, inferences based on confidence regions or tests that fail to satisfy (2) or (3) may be very misleading in finite samples. Of course, as pointed out by Bahadur and Savage (1956), there may be no nontrivial confidence region or test satisfying (2) or (3) when $\mathbf{P}$ is sufficiently rich. For this reason, we will have to restrict $\mathbf{P}$ appropriately in our examples. In the case of confidence regions for or tests about the mean, for instance, we will have to impose a very weak uniform integrability condition. See also Kabaila (1995), Pötscher (2002), Leeb and Pötscher (2006a, 2006b), Pötscher (2009) for related results in more complicated settings, including post-model selection, shrinkage-estimators and ill-posed problems.

Some of our results on subsampling are closely related to results in Andrews and Guggenberger (2010), which were developed independently and at about the same time as our results. See the discussion on page 431 of Andrews and Guggenberger (2010). Our results show that the question of whether subsampling can be used to construct estimators $\hat{c}_n(\alpha)$ satisfying (1) reduces to a single, succinct requirement on the asymptotic relationship between the distribution of $J_n(x, P)$ and $J_b(x, P)$, where b is the subsample size, whereas the results of Andrews and Guggenberger (2010) require the verification of a larger number of conditions. Moreover, we also provide a converse, showing this requirement on the asymptotic relationship between the distribution of $J_n(x, P)$ and $J_b(x, P)$ is also necessary in the sense that, if the requirement fails, then for some nominal coverage level, the uniform coverage statements fail. Thus our results are stated under essentially the weakest possible conditions, yet are verifiable in a large class of examples. On the other hand, the results of Andrews and Guggenberger (2010) further provide a means of calculating the limiting value of $\inf_{P \in \mathbf{P}} P\{\hat{c}_n(\alpha_1) \leq R_n \leq \hat{c}_n(1 - \alpha_2)\}$ in the case where it may not satisfy (1). To the best of our knowledge, our results on the bootstrap are the first to be stated at this level of generality. An important antecedent is Romano (1989), who studies the uniform asymptotic behavior of confidence regions for a univariate cumulative distribution function. See also Mikusheva (2007), who analyzes the uniform asymptotic behavior of some tests that arise in the context of an autoregressive model.

The remainder of the paper is organized as follows. In Section 2, we present the conditions under which $\hat{c}_n(\alpha)$ satisfying (1) may be constructed using subsampling or the bootstrap. We then provide in Section 3 several applications of our general results. These applications include the multivariate mean, testing moment inequalities, multiple testing, the empirical process and U-statistics. The discussion of U-statistics is especially noteworthy because it highlights the fact that the assumptions required for the uniform asymptotic validity of subsampling and the bootstrap may differ. In particular, subsampling may be uniformly asymptotically valid under conditions where, as noted by Bickel and Freedman (1981), the bootstrap fails even to be pointwise asymptotically valid. The application to multiple testing is also noteworthy because, despite the enormous recent literature in this area, our results appear to be the first that provide uniformly asymptotically valid inference. Proofs of the main results (Theorems 2.1 and 2.4) can be found in the Appendix; proofs of all other results can be found in Romano and Shaikh (2012), which contains supplementary material. Many of the intermediate results may be of independent interest, including uniform weak laws of large numbers for U-statistics and V-statistics [Lemmas S.17.3 and S.17.4 in Romano and Shaikh (2012), resp.] as well as the aforementioned Lemma A.1.

2. General results.

2.1. *Subsampling.* Let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$. Denote by $J_n(x, P)$ the distribution of a real-valued root $R_n = R_n(X^{(n)}, P)$ under P. The goal is to construct procedures which are valid uniformly in P. In order to describe the subsampling approach to approximate $J_n(x, P)$, let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$, and define $N_n = \binom{n}{b}$. For $i = 1, \dots, N_n$, denote by $X^{n,(b),i}$ the ith subset of data of size b. Below, we present results for two subsampling-based estimators of $J_n(x, P)$. We first consider the estimator given by

$$L_n(x, P) = \frac{1}{N_n} \sum_{1 \le i \le N_n} I\{R_b(X^{n,(b),i}, P) \le x\}. \tag{4}$$

More generally, we will also consider feasible estimators $\hat{L}_n(x)$ in which R_b is replaced by some estimator $\hat{R}_b$, that is,

$$\hat{L}_n(x) = \frac{1}{N_n} \sum_{1 \le i \le N_n} I\{\hat{R}_b(X^{n,(b),i}) \le x\}. \tag{5}$$

Typically, $\hat{R}_b(\cdot) = R_b(\cdot, \hat{P}_n)$, where $\hat{P}_n$ is the empirical distribution, but this is not assumed below. Even though the estimator of $J_n(x, P)$ defined in (4) is infeasible because of its dependence on P, which is unknown, it is useful both as an intermediate step toward establishing some results for the feasible estimator of $J_n(x, P)$ and, as explained in Remarks 2.2 and 2.3, on its own in the construction of some feasible tests and confidence regions.

THEOREM 2.1. *Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$, and define $L_n(x, P)$ as in (4). Then, the following statements are true:*

(i) *If* $\limsup_{n\to\infty} \sup_{P \in \mathbf{P}} \sup_{x \in \mathbf{R}} \{J_b(x, P) - J_n(x, P)\} \le 0$, *then*

$$\liminf_{n\to\infty} \inf_{P \in \mathbf{P}} P\{L_n^{-1}(\alpha_1, P) \le R_n \le L_n^{-1}(1 - \alpha_2, P)\} \ge 1 - \alpha_1 - \alpha_2 \tag{6}$$

holds for $\alpha_1 = 0$ *and any* $0 \leq \alpha_2 < 1$.

(ii) *If* $\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{x\in\mathbf{R}}\{J_n(x, P) - J_b(x, P)\} \leq 0$, *then* (6) *holds for* $\alpha_2 = 0$ *and any* $0 \leq \alpha_1 < 1$.

(iii) *If* $\lim_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{x\in\mathbf{R}} |J_b(x, P) - J_n(x, P)| = 0$, *then* (6) *holds for any* $\alpha_1 \geq 0$ *and* $\alpha_2 \geq 0$ *satisfying* $0 \leq \alpha_1 + \alpha_2 < 1$.

REMARK 2.1. It is typically easy to deduce from the conclusions of Theorem 2.1 stronger results in which the $\liminf_{n\to\infty}$ and $\geq$ in (6) are replaced by $\lim_{n\to\infty}$ and $=$, respectively. For example, in order to assert that (6) holds with $\liminf_{n\to\infty}$ and $\geq$ replaced by $\lim_{n\to\infty}$ and $=$, respectively, all that is required is that

$$\lim_{n\to\infty} P\{L_n^{-1}(\alpha_1, P) \leq R_n \leq L_n^{-1}(1-\alpha_2, P)\} = 1 - \alpha_1 - \alpha_2$$

for some $P \in \mathbf{P}$. This can be verified using the usual arguments for the pointwise asymptotic validity of subsampling. Indeed, it suffices to show for some $P \in \mathbf{P}$ that $J_n(x, P)$ tends in distribution to a limiting distribution $J(x, P)$ that is continuous at the appropriate quantiles. See Politis, Romano and Wolf (1999) for details.

REMARK 2.2. As mentioned earlier, $L_n(x, P)$ defined in (4) is infeasible because it still depends on P, which is unknown, through $R_b(X^{n,(b),i}, P)$. Even so, Theorem 2.1 may be used without modification to construct feasible confidence regions for a parameter of interest $\theta(P)$ provided that $R_n(X^{(n)}, P)$, and therefore $L_n(x, P)$, depends on P only through $\theta(P)$. If this is the case, then one may simply invert tests of the null hypotheses $\theta(P) = \theta$ for all $\theta \in \Theta$ to construct a confidence region for $\theta(P)$. More concretely, suppose $R_n(X^{(n)}, P) = R_n(X^{(n)}, \theta(P))$ and $L_n(x, P) = L_n(x, \theta(P))$. Whenever we may apply part (i) of Theorem 2.1, we have that

$$C_n = \{\theta \in \Theta : R_n(X^{(n)}, \theta) \leq L_n^{-1}(1-\alpha, \theta)\}$$

satisfies (2). Similar conclusions follow from parts (ii) and (iii) of Theorem 2.1.

REMARK 2.3. It is worth emphasizing that even though Theorem 2.1 is stated for roots, it is, of course, applicable in the special case where $R_n(X^{(n)}, P) = T_n(X^{(n)})$. This is especially useful in the context of hypothesis testing. See Example 3.3 for one such instance.

Next, we provide some results for feasible estimators of $J_n(x, P)$. The first result, Corollary 2.1, handles the case of the most basic root, while Theorem 2.2 applies to more general roots needed for many of our applications.

COROLLARY 2.1. *Suppose* $R_n = R_n(X^{(n)}, P) = \tau_n(\hat{\theta}_n - \theta(P))$, *where* $\{\tau_n \in \mathbf{R} : n \geq 1\}$ *is a sequence of normalizing constants*, $\theta(P)$ *is a real-valued parameter* *of interest and* $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ *is an estimator of* $\theta(P)$. *Let* $b = b_n < n$ *be a sequence of positive integers tending to infinity, but satisfying* $b/n \to 0$, *and define*

$$\hat{L}_n(x) = \frac{1}{N_n} \sum_{1 \le i \le N_n} I\{\tau_b(\hat{\theta}_b(X^{n,(b),i}) - \hat{\theta}_n) \le x\}.$$

Then statements (i)–(iii) *of Theorem* 2.1 *hold when* $L_n^{-1}(\cdot, P)$ *is replaced by* $\frac{\tau_n}{\tau_n + \tau_b}\hat{L}_n^{-1}(\cdot)$.

THEOREM 2.2. *Let* $b = b_n < n$ *be a sequence of positive integers tending to infinity, but satisfying* $b/n \to 0$. *Define* $L_n(x, P)$ *as in* (4) *and* $\hat{L}_n(x)$ *as in* (5). *Suppose for all* $\varepsilon > 0$ *that*

$$\sup_{P \in \mathbf{P}} P\Big\{\sup_{x \in \mathbf{R}} \big|\hat{L}_n(x) - L_n(x, P)\big| > \varepsilon\Big\} \to 0. \tag{7}$$

Then, statements (i)–(iii) *of Theorem* 2.1 *hold when* $L_n^{-1}(\cdot, P)$ *is replaced by* $\hat{L}_n^{-1}(\cdot)$.

As a special case, Theorem 2.2 can be applied to Studentized roots.

COROLLARY 2.2. *Suppose*

$$R_n = R_n(X^{(n)}, P) = \frac{\tau_n(\hat{\theta}_n - \theta(P))}{\hat{\sigma}_n},$$

where $\{\tau_n \in \mathbf{R} : n \ge 1\}$ *is a sequence of normalizing constants,* $\theta(P)$ *is a real-valued parameter of interest, and* $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ *is an estimator of* $\theta(P)$, *and* $\hat{\sigma}_n = \hat{\sigma}_n(X^{(n)}) \ge 0$ *is an estimator of some parameter* $\sigma(P) \ge 0$. *Suppose further that*:

(i) *The family of distributions* $\{J_n(x, P) : n \ge 1, P \in \mathbf{P}\}$ *is tight, and any subsequential limiting distribution is continuous.*

(ii) *For any* $\varepsilon > 0$,

$$\sup_{P \in \mathbf{P}} P\left\{\left|\frac{\hat{\sigma}_n}{\sigma(P)} - 1\right| > \varepsilon\right\} \to 0.$$

Let $b = b_n < n$ *be a sequence of positive integers tending to infinity, but satisfying* $b/n \to 0$ *and* $\tau_b/\tau_n \to 0$. *Define*

$$\hat{L}_n(x) = \frac{1}{N_n} \sum_{1 \le i \le N_n} I\left\{\frac{\tau_b(\hat{\theta}_b(X^{n,(b),i}) - \hat{\theta}_n)}{\hat{\sigma}_b(X^{n,(b),i})} \le x\right\}.$$

Then statements (i)–(iii) *of Theorem* 2.1 *hold when* $L_n^{-1}(\cdot, P)$ *is replaced by* $\hat{L}_n^{-1}(\cdot)$.

REMARK 2.4. One can take $\hat{\sigma}_n = \sigma(P)$ in Corollary 2.2. Since $\sigma(P)$ effectively cancels out from both sides of the inequality in the event $\{R_n \le \hat{L}_n^{-1}(1-\alpha)\}$, such a root actually leads to a computationally feasible construction. However, Corollary 2.2 still applies and shows that we can obtain a positive result without the correction factor $\tau_n/(\tau_n + \tau_b)$ present in Corollary 2.1, provided the conditions of Corollary 2.2 hold. For example, if for some $\sigma(P)$, we have that $\tau_n(\hat{\theta}_n - \theta(P_n))/\sigma(P_n)$ is asymptotically standard normal under any sequence $\{P_n \in \mathbf{P} : n \ge 1\}$, then the conditions hold.

REMARK 2.5. In Corollaries 2.1 and 2.2, it is assumed that the rate of convergence τ_n is known. This assumption may be relaxed using techniques described in Politis, Romano and Wolf (1999).

We conclude this section with a result that establishes a converse for Theorems 2.1 and 2.2.

THEOREM 2.3. *Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$ and define $L_n(x, P)$ as in* (4) *and $\hat{L}_n(x)$ as in* (5). *Then the following statements are true*:

(i) *If* $\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{x\in\mathbf{R}}\{J_b(x, P) - J_n(x, P)\} > 0$, *then* (6) *fails for* $\alpha_1 = 0$ *and some* $0 \le \alpha_2 < 1$.

(ii) *If* $\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{x\in\mathbf{R}}\{J_n(x, P) - J_b(x, P)\} > 0$, *then* (6) *fails for* $\alpha_2 = 0$ *and some* $0 \le \alpha_1 < 1$.

(iii) *If* $\liminf_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{x\in\mathbf{R}} |J_b(x, P) - J_n(x, P)| > 0$, *then* (6) *fails for some* $\alpha_1 \ge 0$ *and* $\alpha_2 \ge 0$ *satisfying* $0 \le \alpha_1 + \alpha_2 < 1$.

If, in addition, (7) *holds for any* $\varepsilon > 0$, *then statements* (i)–(iii) *above hold when* $L_n^{-1}(\cdot, P)$ *is replaced by* $\hat{L}_n^{-1}(\cdot)$.

2.2. *Bootstrap*. As before, let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$. Denote by $J_n(x, P)$ the distribution of a real-valued root $R_n = R_n(X^{(n)}, P)$ under P. The goal remains to construct procedures which are valid uniformly in P. The bootstrap approach is to approximate $J_n(\cdot, P)$ by $J_n(\cdot, \hat{P}_n)$ for some estimator $\hat{P}_n$ of P. Typically, $\hat{P}_n$ is the empirical distribution, but this is not assumed in Theorem 2.4 below. Because $\hat{P}_n$ need not a priori even lie in $\mathbf{P}$, it is necessary to introduce a family $\mathbf{P}'$ in which $\hat{P}_n$ lies (at least with high probability). In order for the bootstrap to succeed, we will require that $\rho(\hat{P}_n, P)$ be small for some function (perhaps a metric) $\rho(\cdot, \cdot)$ defined on $\mathbf{P}' \times \mathbf{P}$. For any given problem in which the theorem is applied, $\mathbf{P}$, $\mathbf{P}'$ and ρ must be specified.

THEOREM 2.4. *Let $\rho(\cdot, \cdot)$ be a function on $\mathbf{P}' \times \mathbf{P}$, and let $\hat{P}_n$ be a* (*random*) *sequence of distributions. Then, the following are true*:

(i) *Suppose* $\limsup_{n\to\infty}\sup_{x\in\mathbf{R}}\{J_n(x,Q_n)-J_n(x,P_n)\}\le 0$ *for any sequences* $\{Q_n\in\mathbf{P}':n\ge 1\}$ *and* $\{P_n\in\mathbf{P}:n\ge 1\}$ *satisfying* $\rho(Q_n,P_n)\to 0$. *If*

$$\rho(\hat{P}_n,P_n)\stackrel{P_n}{\to}0 \quad \text{and} \quad P_n\{\hat{P}_n\in\mathbf{P}'\}\to 1 \tag{8}$$

for any sequence $\{P_n\in\mathbf{P}:n\ge 1\}$, *then*

$$\liminf_{n\to\infty}\inf_{P\in\mathbf{P}} P\{J_n^{-1}(\alpha_1,\hat{P}_n)\le R_n\le J_n^{-1}(1-\alpha_2,\hat{P}_n)\}\ge 1-\alpha_1-\alpha_2 \tag{9}$$

holds for $\alpha_1=0$ *and any* $0\le\alpha_2<1$.

(ii) *Suppose* $\limsup_{n\to\infty}\sup_{x\in\mathbf{R}}\{J_n(x,P_n)-J_n(x,Q_n)\}\le 0$ *for any sequences* $\{Q_n\in\mathbf{P}':n\ge 1\}$ *and* $\{P_n\in\mathbf{P}:n\ge 1\}$ *satisfying* $\rho(Q_n,P_n)\to 0$. *If* (8) *holds for any sequence* $\{P_n\in\mathbf{P}:n\ge 1\}$, *then* (9) *holds for* $\alpha_2=0$ *and any* $0\le\alpha_1<1$.

(iii) *Suppose* $\lim_{n\to\infty}\sup_{x\in\mathbf{R}}|J_n(x,Q_n)-J_n(x,P_n)|=0$ *for any sequences* $\{Q_n\in\mathbf{P}':n\ge 1\}$ *and* $\{P_n\in\mathbf{P}:n\ge 1\}$ *satisfying* $\rho(Q_n,P_n)\to 0$. *If* (8) *holds for any sequence* $\{P_n\in\mathbf{P}:n\ge 1\}$, *then* (9) *holds for any* $\alpha_1\ge 0$ *and* $\alpha_2\ge 0$ *satisfying* $0\le\alpha_1+\alpha_2<1$.

REMARK 2.6. It is typically easy to deduce from the conclusions of Theorem 2.4 stronger results in which the $\liminf_{n\to\infty}$ and $\ge$ in (9) are replaced by $\lim_{n\to\infty}$ and $=$, respectively. For example, in order to assert that (9) holds with $\liminf_{n\to\infty}$ and $\ge$ replaced by $\lim_{n\to\infty}$ and $=$, respectively, all that is required is that

$$\lim_{n\to\infty} P\{J_n^{-1}(\alpha_1,\hat{P}_n)\le R_n\le J_n^{-1}(1-\alpha_2,\hat{P}_n)\}=1-\alpha_1-\alpha_2$$

for some $P\in\mathbf{P}$. This can be verified using the usual arguments for the pointwise asymptotic validity of the bootstrap. See Politis, Romano and Wolf (1999) for details.

REMARK 2.7. In some cases, it is possible to construct estimators $\hat{J}_n(x)$ of $J_n(x,P)$ that are uniformly consistent over a large class of distributions $\mathbf{P}$ in the sense that for any $\varepsilon>0$

$$\sup_{P\in\mathbf{P}} P\{\rho(\hat{J}_n(\cdot),J_n(\cdot,P))>\varepsilon\}\to 0, \tag{10}$$

where ρ is the Levy metric or some other metric compatible with the weak topology. Yet a result such as (10) is not strong enough to yield uniform coverage statements such as those in Theorems 2.1 and 2.4. In other words, such conclusions do not follow from uniform approximations of the distribution of interest if the quality of the approximation is measured in terms of metrics metrizing weak convergence. To see this, consider the following simple example.

EXAMPLE 2.1. Let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P_\theta = \text{Bernoulli}(\theta)$. Denote by $J_n(x, P_\theta)$ the distribution of the root $R_n = \sqrt{n}(\hat{\theta}_n - \theta)$ under P_θ, where $\hat{\theta}_n = \bar{X}_n$. Let $\hat{P}_n$ be the empirical distribution of $X^{(n)}$ or, equivalently, $P_{\hat{\theta}_n}$. Lemma S.1.1 in Romano and Shaikh (2012) implies for any $\varepsilon > 0$ that

$$\sup_{0 \le \theta \le 1} P_\theta\{\rho(J_n(\cdot, \hat{P}_n), J_n(\cdot, P_\theta)) > \varepsilon\} \to 0, \tag{11}$$

whenever ρ is a metric compatible with the weak topology. Nevertheless, it follows from the argument on page 78 of Romano (1989) that the coverage statements in Theorem 2.4 fail to hold provided that both α_1 and α_2 do not equal zero. Indeed, consider part (i) of Theorem 2.4. Suppose $\alpha_1 = 0$ and $0 < \alpha_2 < 1$. For a given n and $\delta > 0$, let $\theta_n = (1 - \delta)^{1/n}$. Under P_{θ_n}, the event $X_1 = \cdots = X_n = 1$ has probability $1 - \delta$. Moreover, whenever such an event occurs, $R_n > J_n^{-1}(1 - \alpha_2, \hat{P}_n) = 0$. Therefore, $P_{\theta_n}\{J_n^{-1}(\alpha_1, \hat{P}_n) \le R_n \le J_n^{-1}(1 - \alpha_2, \hat{P}_n)\} \le \delta$. Since the choice of δ was arbitrary, it follows that

$$\liminf_{n\to\infty} \inf_{0 \le \theta \le 1} P_\theta\{J_n^{-1}(\alpha_1, \hat{P}_n) \le R_n \le J_n^{-1}(1 - \alpha_2, \hat{P}_n)\} = 0.$$

A similar argument establishes the result for parts (ii) and (iii) of Theorem 2.4.

On the other hand, when ρ is the Kolmogorov metric, (11) holds when the supremum over $0 \le \theta \le 1$ is replaced with a supremum over $\delta < \theta < 1 - \delta$ for some $\delta > 0$. Moreover, when θ is restricted to such an interval, the coverage statements in Theorem 2.4 hold as well.

3. Applications. Before proceeding, it is useful to introduce some notation that will be used frequently throughout many of the examples below. For a distribution P on $\mathbf{R}^k$, denote by $\mu(P)$ the mean of P, by $\Sigma(P)$ the covariance matrix of P, and by $\Omega(P)$ the correlation matrix of P. For $1 \le j \le k$, denote by $\mu_j(P)$ the jth component of $\mu(P)$ and by $\sigma_j^2(P)$ the jth diagonal element of $\Sigma(P)$. In all of our examples, $X^{(n)} = (X_1, \dots, X_n)$ will be an i.i.d. sequence of random variables with distribution P and $\hat{P}_n$ will denote the empirical distribution of $X^{(n)}$. As usual, we will denote by $\bar{X}_n = \mu(\hat{P}_n)$ the usual sample mean, by $\hat{\Sigma}_n = \Sigma(\hat{P}_n)$ the usual sample covariance matrix and by $\hat{\Omega}_n = \Omega(\hat{P}_n)$ the usual sample correlation matrix. For $1 \le j \le k$, denote by $\bar{X}_{j,n}$ the jth component of $\bar{X}_n$ and by $S_{j,n}^2$ the jth diagonal element of $\hat{\Sigma}_n$. Finally, we say that a family of distributions $\mathbf{Q}$ on the real line satisfies the standardized uniform integrability condition if

$$\lim_{\lambda\to\infty} \sup_{Q \in \mathbf{Q}} E_Q\left[\left(\frac{Y - \mu(Q)}{\sigma(Q)}\right)^2 I\left\{\left|\frac{Y - \mu(Q)}{\sigma(Q)}\right| > \lambda\right\}\right] = 0. \tag{12}$$

In the preceding expression, Y denotes a random variable with distribution Q. The use of the term standardized to describe (12) reflects that fact that the variable Y is centered around its mean and normalized by its standard deviation.

3.1. *Subsampling*.

EXAMPLE 3.1 (Multivariate nonparametric mean). Let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}^k$. Suppose one wishes to construct a rectangular confidence region for $\mu(P)$. For this purpose, a natural choice of root is

$$R_n(X^{(n)}, P) = \max_{1 \le j \le k} \frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P))}{S_{j,n}}. \tag{13}$$

In this setup, we have the following theorem:

THEOREM 3.1. *Denote by $\mathbf{P}_j$ the set of distributions formed from the jth marginal distributions of the distributions in $\mathbf{P}$. Suppose $\mathbf{P}$ is such that (12) is satisfied with $\mathbf{Q} = \mathbf{P}_j$ for all $1 \le j \le k$. Let $J_n(x, P)$ be the distribution of the root (13). Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$ and define $L_n(x, P)$ by (4). Then*

$$\begin{aligned} &\lim_{n\to\infty} \inf_{P \in \mathbf{P}} P\left\{ L_n^{-1}(\alpha_1, P) \le \max_{1 \le j \le k} \frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P))}{S_{j,n}} \le L_n^{-1}(1 - \alpha_2, P) \right\} \\ &\quad = 1 - \alpha_1 - \alpha_2 \end{aligned} \tag{14}$$

for any $\alpha_1 \ge 0$ and $\alpha_2 \ge 0$ such that $0 \le \alpha_1 + \alpha_2 < 1$. Furthermore, (14) remains true if $L_n^{-1}(\cdot, P)$ is replaced by $\hat{L}_n^{-1}(\cdot)$, where $\hat{L}_n(x)$ is defined by (5) with $\hat{R}_b(X^{n,(b),i}) = R_b(X^{n,(b),i}, \hat{P}_n)$.

Under suitable restrictions, Theorem 3.1 generalizes to the case where the root is given by

$$R_n(X^{(n)}, P) = f(Z_n(P), \hat{\Omega}_n), \tag{15}$$

where f is a continuous, real-valued function and

$$Z_n(P) = \left(\frac{\sqrt{n}(\bar{X}_{1,n} - \mu_1(P))}{S_{1,n}}, \dots, \frac{\sqrt{n}(\bar{X}_{k,n} - \mu_k(P))}{S_{k,n}} \right)'. \tag{16}$$

In particular, we have the following theorem:

THEOREM 3.2. *Let $\mathbf{P}$ be defined as in Theorem 3.1. Let $J_n(x, P)$ be the distribution of root (15), where f is continuous.*

(i) *Suppose further that for all $x \in \mathbf{R}$ that*

$$P_n\{f(Z_n(P_n), \Omega(\hat{P}_n)) \le x\} \to P\{f(Z, \Omega) \le x\}, \tag{17}$$

$$P_n\{f(Z_n(P_n), \Omega(\hat{P}_n)) < x\} \to P\{f(Z, \Omega) < x\} \tag{18}$$

for any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$ *such that* $Z_n(P_n) \stackrel{d}{\to} Z$ *under* P_n *and* $\Omega(\hat{P}_n) \stackrel{P_n}{\to} \Omega$, *where* $Z \sim N(0, \Omega)$. *Then*

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P\{L_n^{-1}(\alpha_1, P) \leq f(Z_n(P), \hat{\Omega}_n) \leq L_n^{-1}(1-\alpha_2, P)\} \geq 1 - \alpha_1 - \alpha_2 \tag{19}$$

for any $\alpha_1 \geq 0$ *and* $\alpha_2 \geq 0$ *such that* $0 \leq \alpha_1 + \alpha_2 < 1$.

(ii) *Suppose further that if* $Z \sim N(0, \Omega)$ *for some* Ω *satisfying* $\Omega_{j,j} = 1$ *for all* $1 \leq j \leq k$, *then* $f(Z, \Omega)$ *is continuously distributed. Then,* (19) *remains true if* $L_n^{-1}(\cdot, P)$ *is replaced by* $\hat{L}_n^{-1}(\cdot)$, *where* $\hat{L}_n(x)$ *is defined by* (5) *with* $\hat{R}_b(X^{n,(b),i}) = R_b(X^{n,(b),i}, \hat{P}_n)$. *Moreover, the* $\liminf_{n\to\infty}$ *and* $\geq$ *may be replaced by* $\lim_{n\to\infty}$ *and* $=$, *respectively.*

In order to verify (17) and (18) in Theorem 3.2, it suffices to assume that $f(Z, \Omega)$ is continuously distributed. Under the assumptions of the theorem, however, $f(Z, \Omega)$ need not be continuously distributed. In this case, (17) and (18) hold immediately for any x at which $P\{(Z, \Omega) \leq x\}$ is continuous, but require a further argument for x at which $P\{(Z, \Omega) \leq x\}$ is discontinuous. See, for example, the proof of Theorem 3.9, which relies on Theorem 3.8, where the same requirement appears.

EXAMPLE 3.2 (Constrained univariate nonparametric mean). Andrews (2000) considers the following example. Let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}$. Suppose it is known that $\mu(P) \geq 0$ for all $P \in \mathbf{P}$ and one wishes to construct a confidence interval for $\mu(P)$. A natural choice of root in this case is

$$R_n = R_n(X^{(n)}, P) = \sqrt{n}(\max\{\bar{X}_n, 0\} - \mu(P)).$$

This root differs from the one considered in Theorem 3.1 and the ones discussed in Theorem 3.2 in the sense that under weak assumptions on $\mathbf{P}$,

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{x\in\mathbf{R}} \{J_b(x, P) - J_n(x, P)\} \leq 0 \tag{20}$$

holds, but

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{x\in\mathbf{R}} \{J_n(x, P) - J_b(x, P)\} \leq 0 \tag{21}$$

fails to hold. To see this, suppose (12) holds with $\mathbf{Q} = \mathbf{P}$. Note that

$$J_b(x, P) = P\{\max\{Z_b(P), -\sqrt{b}\mu(P)\} \leq x\},$$
$$J_n(x, P) = P\{\max\{Z_n(P), -\sqrt{n}\mu(P)\} \leq x\},$$

where $Z_b(P) = \sqrt{b}(\bar{X}_b - \mu(P))$ and $Z_n(P) = \sqrt{n}(\bar{X}_n - \mu(P))$. Since $\sqrt{b}\mu(P) \leq \sqrt{n}\mu(P)$ for any $P \in \mathbf{P}$, $J_b(x, P) - J_n(x, P)$ is bounded from above by

$$P\{\max\{Z_b(P), -\sqrt{n}\mu(P)\} \leq x\} - J_n(x, P).$$

It now follows from the uniform central limit theorem established by Lemma 3.3.1 of Romano and Shaikh (2008) and Theorem 2.11 of Bhattacharya and Ranga Rao (1976) that (20) holds. It therefore follows from Theorem 2.1 that (6) holds with $\alpha_1 = 0$ and any $0 \leq \alpha_2 < 1$. To see that (21) fails, suppose further that $\{Q_n : n \geq 1\} \subseteq \mathbf{P}$, where $Q_n = N(h/\sqrt{n}, 1)$ for some $h > 0$. For $Z \sim N(0, 1)$,

$$J_n(x, Q_n) = P\{\max(Z, -h) \leq x\},$$
$$J_b(x, Q_n) = P\{\max(Z, -h\sqrt{b}/\sqrt{n}) \leq x\}.$$

The left-hand side of (21) is therefore greater than or equal to

$$\limsup_{n\to\infty}\left(P\{\max(Z, -h) \leq x\} - P\{\max(Z, -h\sqrt{b}/\sqrt{n}) \leq x\}\right)$$

for any x. In particular, if $-h < x < 0$, then the second term is zero for large enough n, and so the limiting value is $P\{Z \leq x\} = \Phi(x) > 0$. It therefore follows from Theorem 2.3 that (6) fails for $\alpha_2 = 0$ and some $0 \leq \alpha_1 < 1$. On the other hand, (6) holds with $\alpha_2 = 0$ and any $0.5 < \alpha_1 < 1$. To see this, consider any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$ and the event $\{L_n^{-1}(\alpha_1, P_n) \leq R_n\}$. For the root in this example, this event is scale invariant. So, in calculating the probability of this event, we may without loss of generality assume $\sigma^2(P_n) = 1$. Since $\mu(P_n) \geq 0$, we have for any $x \geq 0$ that

$$J_n(x, P_n) = P\{\max\{Z_n(P_n), -\sqrt{n}\mu(P_n)\} \leq x\} = P\{Z_n(P_n) \leq x\} \to \Phi(x)$$

and similarly for $J_b(x, P_n)$. Using the usual subsampling arguments, it is thus possible to show for $0.5 < \alpha_1 < 1$ that

$$L_n^{-1}(\alpha_1, P_n) \overset{P_n}{\to} \Phi^{-1}(\alpha_1).$$

The desired conclusion therefore follows from Slutsky's theorem. Arguing as the the proof of Corollary 2.2 and Remark 2.4, it can be shown that the same results hold when $L_n^{-1}(\cdot, P)$ is replaced by $\hat{L}_n^{-1}(\cdot)$, where $\hat{L}_n(x)$ is defined as $L_n(x, P)$ is defined but with $\mu(P)$ replaced by $\bar{X}_n$.

EXAMPLE 3.3 (Moment inequalities). The generality of Theorem 2.1 illustrated in Example 3.2 is also useful when testing multisided hypotheses about the mean. To see this, let $X^{(n)} = (X_1, \ldots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}^k$. Define $\mathbf{P}_0 = \{P \in \mathbf{P} : \mu(P) \leq 0\}$ and $\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0$. Consider testing the null hypothesis that $P \in \mathbf{P}_0$ versus the alternative hypothesis that $P \in \mathbf{P}_1$ at level $\alpha \in (0, 1)$. Such hypothesis testing problems have recently received considerable attention in the "moment inequality" literature in econometrics. See, for example, Andrews and Soares (2010), Andrews and Guggenberger (2010), Andrews and Barwick (2012), Bugni (2010), Canay (2010) and Romano and Shaikh (2008, 2010). Theorem 2.1 may be used to construct tests that are uniformly consistent in level in the sense that (3) holds under weak assumptions on $\mathbf{P}$. Formally, we have the following theorem:

THEOREM 3.3. *Let* $\mathbf{P}$ *be defined as in Theorem* 3.1. *Let* $J_n(x, P)$ *be the distribution of*

$$T_n(X^{(n)}) = \max_{1 \le j \le k} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}}.$$

Let $b = b_n < n$ *be a sequence of positive integers tending to infinity, but satisfying* $b/n \to 0$ *and define* $L_n(x)$ *by the right-hand side of* (4) *with* $R_n(X^{(n)}, P) = T_n(X^{(n)})$. *Then, the test defined by*

$$\phi_n(X^{(n)}) = I\{T_n(X^{(n)}) > L_n^{-1}(1-\alpha)\}$$

satisfies (3) *for any* $0 < \alpha < 1$.

The argument used to establish Theorem 3.3 is essentially the same as the one presented in Romano and Shaikh (2008) for

$$T_n(X^{(n)}) = \sum_{1 \le j \le k} \max\{\sqrt{n}\bar{X}_{j,n}, 0\}^2,$$

though Lemma S.6.1 in Romano and Shaikh (2012) is needed for establishing (20) here because of Studentization. Related results are obtained by Andrews and Guggenberger (2009).

EXAMPLE 3.4 (Multiple testing). We now illustrate the use of Theorem 2.1 to construct tests of multiple hypotheses that behave well uniformly over a large class of distributions. Let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}^k$, and consider testing the family of null hypotheses

$$H_j : \mu_j(P) \le 0 \qquad \text{for } 1 \le j \le k \tag{22}$$

versus the alternative hypotheses

$$H_j' : \mu_j(P) > 0 \qquad \text{for } 1 \le j \le k \tag{23}$$

in a way that controls the familywise error rate at level $0 < \alpha < 1$ in the sense that

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}} \mathrm{FWER}_P \le \alpha, \tag{24}$$

where

$$\mathrm{FWER}_P = P\{\text{reject some } H_j \text{ with } \mu_j(P) \le 0\}.$$

For $K \subseteq \{1, \dots, k\}$, define $L_n(x, K)$ according to the right-hand side of (4) with

$$R_n\big(X^{(n)}, P\big) = \max_{j \in K} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}},$$

and consider the following stepwise multiple testing procedure:

ALGORITHM 3.1. Step 1: Set $K_1 = \{1, \dots, k\}$. If

$$\max_{j \in K_1} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} \le L_n^{-1}(1 - \alpha, K_1),$$

then stop. Otherwise, reject any H_j with

$$\frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} > L_n^{-1}(1 - \alpha, K_1)$$

and continue to Step 2 with

$$K_2 = \left\{ j \in K_1 : \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} \le L_n^{-1}(1 - \alpha, K_1) \right\}.$$

$\vdots$

Step s: If

$$\max_{j \in K_s} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} \le L_n^{-1}(1 - \alpha, K_s),$$

then stop. Otherwise, reject any H_j with

$$\frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} > L_n^{-1}(1 - \alpha, K_s)$$

and continue to Step $s + 1$ with

$$K_{s+1} = \left\{ j \in K_s : \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} \le L_n^{-1}(1 - \alpha, K_s) \right\}.$$

$\vdots$

We have the following theorem:

THEOREM 3.4. *Let* $\mathbf{P}$ *be defined as in Theorem* 3.1. *Let* $b = b_n < n$ *be a sequence of positive integers tending to infinity, but satisfying* $b/n \to 0$. *Then, Algorithm* 3.1 *satisfies*

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}} \mathrm{FWER}_P \le \alpha \tag{25}$$

for any $0 < \alpha < 1$.

It is, of course, possible to extend the analysis in a straightforward way to two-sided testing. See also Romano and Shaikh (2010) for related results about a multiple testing problem involving an infinite number of null hypotheses.

EXAMPLE 3.5 (Empirical process on **R**). Let $X^{(n)} = (X_1, \ldots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on **R**. Suppose one wishes to construct a confidence region for the cumulative distribution function associated with P, that is, $P\{(-\infty, t]\}$. For this purpose a natural choice of root is

$$\sup_{t \in \mathbf{R}} \sqrt{n} \big| \hat{P}_n\{(-\infty, t]\} - P\{(-\infty, t]\} \big|. \tag{26}$$

In this setting, we have the following theorem:

THEOREM 3.5. *Fix any $\varepsilon \in (0, 1)$, and let*

$$\mathbf{P} = \big\{ P \text{ on } \mathbf{R} : \varepsilon < P\big\{(-\infty, t]\big\} < 1 - \varepsilon \text{ for some } t \in \mathbf{R} \big\}. \tag{27}$$

Let $J_n(x, P)$ be the distribution of root (26). Then

$$\begin{aligned} \lim_{n \to \infty} \inf_{P \in \mathbf{P}} P\Big\{ L_n^{-1}(\alpha_1, P) \le \sup_{t \in \mathbf{R}} \sqrt{n} \big| \hat{P}_n\{(-\infty, t]\} - P\{(-\infty, t]\} \big| \\ \le L_n^{-1}(1 - \alpha_2, P) \Big\} \\ = 1 - \alpha_1 - \alpha_2 \end{aligned} \tag{28}$$

for any $\alpha_1 \ge 0$ and $\alpha_2 \ge 0$ such that $0 \le \alpha_1 + \alpha_2 < 1$. Furthermore, (28) remains true if $L_n^{-1}(\cdot, P)$ is replaced by $\hat{L}_n^{-1}(\cdot)$, where $\hat{L}_n(x)$ is defined by (5) with $\hat{R}_b(X^{n,(b),i}) = R_b(X^{n,(b),i}, \hat{P}_n)$.

EXAMPLE 3.6 (One sample U-statistics). Let $X^{(n)} = (X_1, \ldots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on **R**. Suppose one wishes to construct a confidence region for

$$\theta(P) = \theta_h(P) = E_P\big[h(X_1, \ldots, X_m)\big], \tag{29}$$

where h is a symmetric kernel of degree m. The usual estimator of $\theta(P)$ in this case is given by the U-statistic

$$\hat{\theta}_n = \hat{\theta}_n\big(X^{(n)}\big) = \frac{1}{\binom{n}{m}} \sum_c h(X_{i_1}, \ldots, X_{i_m}).$$

Here, $\sum_c$ denotes summation over all $\binom{n}{m}$ subsets $\{i_1, \ldots, i_m\}$ of $\{1, \ldots, n\}$. A natural choice of root is therefore given by

$$R_n\big(X^{(n)}, P\big) = \sqrt{n}\big(\hat{\theta}_n - \theta(P)\big). \tag{30}$$

In this setting, we have the following theorem:

THEOREM 3.6. *Let*

$$g(x,P) = g_h(x,P) = E_P\big[h(x, X_2, \dots, X_m)\big] - \theta(P) \tag{31}$$

and

$$\sigma_h^2(P) = m^2 \operatorname{Var}_P\big[g(X_i, P)\big]. \tag{32}$$

Suppose **P** *satisfies the uniform integrability condition*

$$\lim_{\lambda\to\infty} \sup_{P\in\mathbf{P}} E_P\left[\frac{g^2(X_i,P)}{\sigma_h^2(P)} I\left\{\left|\frac{g(X_i,P)}{\sigma_h(P)}\right| > \lambda\right\}\right] = 0 \tag{33}$$

and

$$\sup_{P\in\mathbf{P}} \frac{\operatorname{Var}_P[h(X_1,\dots,X_m)]}{\sigma^2(P)} < \infty. \tag{34}$$

Let $J_n(x,P)$ be the distribution of the root (30). *Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$, and define $L_n(x,P)$ by* (4). *Then*

$$\begin{aligned} &\lim_{n\to\infty} \inf_{P\in\mathbf{P}} P\big\{L_n^{-1}(\alpha_1, P) \le \sqrt{n}\big(\hat\theta_n - \theta(P)\big) \le L_n^{-1}(1-\alpha_2, P)\big\} \\ &\quad = 1 - \alpha_1 - \alpha_2 \end{aligned} \tag{35}$$

for any $\alpha_1 \ge 0$ and $\alpha_2 \ge 0$ such that $0 \le \alpha_1 + \alpha_2 < 1$. Furthermore, (35) *remains true if $L_n^{-1}(\cdot, P)$ is replaced by $\hat L_n^{-1}(\cdot)$, where $\hat L_n(x)$ is defined by* (5) *with $\hat R_b(X^{n,(b),i}) = R_b(X^{n,(b),i}, \hat P_n)$.*

3.2. *Bootstrap.*

EXAMPLE 3.7 (Multivariate nonparametric mean). Let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}^k$. Suppose one wishes to construct a rectangular confidence region for $\mu(P)$. As described in Example 3.1, a natural choice of root in this case is given by (13). In this setting, we have the following theorem, which is a bootstrap counterpart to Theorem 3.1:

THEOREM 3.7. *Let* **P** *be defined as in Theorem* 3.1. *Let $J_n(x,P)$ be the distribution of the root* (13). *Then*

$$\begin{aligned} &\lim_{n\to\infty} \inf_{P\in\mathbf{P}} P\left\{J_n^{-1}(\alpha_1, \hat P_n) \le \max_{1\le j\le k} \frac{\sqrt{n}(\bar X_{j,n} - \mu_j(P))}{S_{j,n}} \le J_n^{-1}(1-\alpha_2, \hat P_n)\right\} \\ &\quad = 1 - \alpha_1 - \alpha_2 \end{aligned} \tag{36}$$

for any $\alpha_1 \ge 0$ and $\alpha_2 \ge 0$ such that $0 \le \alpha_1 + \alpha_2 < 1$.

Theorem 3.7 generalizes in the same way that Theorem 3.1 generalizes. In particular, we have the following result:

THEOREM 3.8. *Let* $\mathbf{P}$ *be defined as in Theorem* 3.1. *Let* $J_n(x, P)$ *be the distribution of the root* (15). *Suppose* f *is continuous. Suppose further that for all* $x \in \mathbf{R}$

$$P_n\{f(Z_n(P_n), \Omega(\hat{P}_n)) \le x\} \to P\{f(Z, \Omega) \le x\}, \tag{37}$$

$$P_n\{f(Z_n(P_n), \Omega(\hat{P}_n)) < x\} \to P\{f(Z, \Omega) < x\} \tag{38}$$

for any sequence $\{P_n \in \mathbf{P} : n \ge 1\}$ *such that* $Z_n(P_n) \xrightarrow{d} Z$ *under* P_n *and* $\Omega(\hat{P}_n) \xrightarrow{P_n} \Omega$, *where* $Z \sim N(0, \Omega)$. *Then*

$$\begin{aligned} &\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P\{J_n^{-1}(\alpha_1, \hat{P}_n) \le f(Z_n(P), \hat{\Omega}_n) \le J_n^{-1}(1-\alpha_2, \hat{P}_n)\} \\ &\quad \ge 1 - \alpha_1 - \alpha_2 \end{aligned} \tag{39}$$

for any $\alpha_1 \ge 0$ *and* $\alpha_2 \ge 0$ *such that* $0 \le \alpha_1 + \alpha_2 < 1$.

EXAMPLE 3.8 (Moment inequalities). Let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}^k$ and define $\mathbf{P}_0$ and $\mathbf{P}_1$ as in Example 3.3. Andrews and Barwick (2012) propose testing the null hypothesis that $P \in \mathbf{P}_0$ versus the alternative hypothesis that $P \in \mathbf{P}_1$ at level $\alpha \in (0, 1)$ using an "adjusted quasi-likelihood ratio" statistic $T_n(X^{(n)})$ defined as follows:

$$T_n(X^{(n)}) = \inf_{t\in\mathbf{R}^k : t\le 0} W_n(t)'\tilde{\Omega}_n^{-1} W_n(t).$$

Here, $t \le 0$ is understood to mean that the inequality holds component-wise,

$$W_n(t) = \left(\frac{\sqrt{n}(\bar{X}_{1,n} - t_1)}{S_{1,n}}, \dots, \frac{\sqrt{n}(\bar{X}_{k,n} - t_k)}{S_{k,n}}\right)'$$

and

$$\tilde{\Omega}_n = \max\{\varepsilon - \det(\hat{\Omega}_n), 0\} I_k + \hat{\Omega}_n, \tag{40}$$

where $\varepsilon > 0$ and I_k is the k-dimensional identity matrix. Andrews and Barwick (2012) propose a procedure for constructing critical values for $T_n(X^{(n)})$ that they term "refined moment selection." For illustrative purposes, we instead consider in the following theorem a simpler construction.

THEOREM 3.9. *Let* $\mathbf{P}$ *be defined as in Theorem* 3.1. *Let* $J_n(x, P)$ *be the distribution of the root*

$$R_n(X^{(n)}, P) = \inf_{t\in\mathbf{R}^k : t\le 0} (Z_n(P) - t)'\tilde{\Omega}_n^{-1}(Z_n(P) - t), \tag{41}$$

where $Z_n(P)$ is defined as in (16). *Then*, *the test defined by*

$$\phi_n(X^{(n)}) = I\{T_n(X^{(n)}) > J_n^{-1}(1-\alpha, \hat{P}_n)\}$$

satisfies (3) *for any* $0 < \alpha < 1$.

Theorem 3.9 generalizes in a straightforward fashion to other choices of test statistics, including the one used in Theorem 3.3. On the other hand, even when the underlying choice of test statistic is the same, the first-order asymptotic properties of the tests in Theorems 3.9 and 3.3 will differ. For other ways of constructing critical values that are more similar to the construction given in Andrews and Barwick (2012), see Romano, Shaikh and Wolf (2012).

EXAMPLE 3.9 (Multiple testing). Theorem 2.4 may be used in the same way that Theorem 2.1 was used in Example 3.4 to construct tests of multiple hypotheses that behave well uniformly over a large class of distributions. To see this, let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}^k$, and again consider testing the family of null hypotheses (22) versus the alternative hypotheses (23) in a way that satisfies (24) for $\alpha \in (0, 1)$. For $K \subseteq \{1, \dots, k\}$, let $J_n(x, K, P)$ be the distribution of the root

$$R_n(X^{(n)}, P) = \max_{j \in K} \frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P))}{S_{j,n}}$$

under P, and consider the stepwise multiple testing procedure given by Algorithm 3.1 with $L_n^{-1}(1-\alpha, K_j)$ replaced by $J_n^{-1}(1-\alpha, K_j, \hat{P}_n)$. We have the following theorem, which is a bootstrap counterpart to Theorem 3.4:

THEOREM 3.10. *Let* $\mathbf{P}$ *be defined as in Theorem* 3.1. *Then Algorithm* 3.1 *with* $L_n^{-1}(1-\alpha, K_j)$ *replaced by* $J_n^{-1}(1-\alpha, K_j, \hat{P}_n)$ *satisfies* (25) *for any* $0 < \alpha < 1$.

It is, of course, possible to extend the analysis in a straightforward way to two-sided testing.

EXAMPLE 3.10 (Empirical process on $\mathbf{R}$). Let $X^{(n)} = (X_1, \dots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}$. Suppose one wishes to construct a confidence region for the cumulative distribution function associated with P, that is, $P\{(-\infty, t]\}$. As described in Example 3.5, a natural choice of root in this case is given by (26). In this setting, we have the following theorem, which is a bootstrap counterpart to Theorem 3.5:

THEOREM 3.11. *Fix any* $\varepsilon \in (0, 1)$, *and let* $\mathbf{P}$ *be defined as in Theorem* 3.5. *Let* $J_n(x, P)$ *be the distribution of the root* (26). *Denote by* $\hat{P}_n$ *the empirical dis-*

tribution of $X^{(n)}$. *Then*

$$\lim_{n\to\infty}\inf_{P\in\mathbf{P}} P\Big\{J_n^{-1}(\alpha_1,\hat{P}_n)\le \sup_{t\in\mathbf{R}}\sqrt{n}\big|\hat{P}_n\{(-\infty,t]\}-P\{(-\infty,t]\}\big| \le J_n^{-1}(1-\alpha_2,\hat{P}_n)\Big\} = 1-\alpha_1-\alpha_2$$

for any $\alpha_1\ge 0$ *and* $\alpha_2\ge 0$ *such that* $0\le\alpha_1+\alpha_2<1$.

Some of the conclusions of Theorem 3.11 can be found in Romano (1989), though the method of proof given in Romano and Shaikh (2012) is quite different.

EXAMPLE 3.11 (One sample U-statistics). Let $X^{(n)}=(X_1,\ldots,X_n)$ be an i.i.d. sequence of random variables with distribution $P\in\mathbf{P}$ on $\mathbf{R}$ and let h be a symmetric kernel of degree m. Suppose one wishes to construct a confidence region for $\theta(P)=\theta_h(P)$ given by (29). As described in Example 3.6, a natural choice of root in this case is given by (30). Before proceeding, it is useful to introduce the following notation. For an arbitrary kernel $\tilde{h}$, $\varepsilon>0$ and $B>0$, denote by $\mathbf{P}_{\tilde{h},\varepsilon,B}$ the set of all distributions P on $\mathbf{R}$ such that

$$E_P\big[\big|\tilde{h}(X_1,\ldots,X_m)-\theta_{\tilde{h}}(P)\big|^{\varepsilon}\big]\le B. \tag{42}$$

Similarly, for an arbitrary kernel $\tilde{h}$ and $\delta>0$, denote by $\mathbf{S}_{\tilde{h},\delta}$ the set of all distributions P on $\mathbf{R}$ such that

$$\sigma^2_{\tilde{h}}(P)\ge\delta, \tag{43}$$

where $\sigma^2_{\tilde{h}}(P)$ is defined as in (32). Finally, for an arbitrary kernel $\tilde{h}$, $\varepsilon>0$ and $B>0$, let $\bar{\mathbf{P}}_{\tilde{h},\varepsilon,B}$ be the set of distributions P on $\mathbf{R}$ such that

$$E_P\big[\big|\tilde{h}(X_{i_1},\ldots,X_{i_m})-\theta_{\tilde{h}}(P)\big|^{\varepsilon}\big]\le B,$$

whenever $1\le i_j\le n$ for all $1\le j\le m$. Using this notation, we have the following theorem:

THEOREM 3.12. *Define the kernel* h' *of degree* $2m$ *according to the rule*

$$\begin{aligned} h'(x_1,\ldots,x_{2m}) &= h(x_1,\ldots,x_m)h(x_1,x_{m+2},\ldots,x_{2m}) \\ &\quad - h(x_1,\ldots,x_m)h(x_{m+1},\ldots,x_{2m}). \end{aligned} \tag{44}$$

Suppose

$$\mathbf{P}\subseteq\mathbf{P}_{h,2+\delta,B}\cap\mathbf{S}_{h,\delta}\cap\bar{\mathbf{P}}_{h',1+\delta,B}\cap\bar{\mathbf{P}}_{h,2+\delta,B}$$

for some $\delta > 0$ *and* $B > 0$. *Let* $J_n(x, P)$ *be the distribution of the root* R_n *defined by* (30). *Then*

$$\lim_{n\to\infty} \inf_{P\in\mathbf{P}} P\{J_n^{-1}(\alpha_1, \hat{P}_n) \le \sqrt{n}(\hat{\theta}_n - \theta(P)) \le J_n^{-1}(1-\alpha_2, \hat{P}_n)\} = 1 - \alpha_1 - \alpha_2$$

for any α_1 *and* α_2 *such that* $0 \le \alpha_1 + \alpha_2 < 1$.

Note that the kernel h' defined in (44) arises in the analysis of the estimated variance of the U-statistic. Note further that the conditions on $\mathbf{P}$ in Theorem 3.12 are stronger than the conditions on $\mathbf{P}$ in Theorem 3.6. While it may be possible to weaken the restrictions on $\mathbf{P}$ in Theorem 3.12 some, it is not possible to establish the conclusions of Theorem 3.12 under the conditions on $\mathbf{P}$ in Theorem 3.6. Indeed, as shown by Bickel and Freedman (1981), the bootstrap based on the root R_n defined by (30) need not be even pointwise asymptotically valid under the conditions on $\mathbf{P}$ in Theorem 3.6.

APPENDIX

A.1. Proof of Theorem 2.1.

LEMMA A.1. *If* F *and* G *are* (*nonrandom*) *distribution functions on* $\mathbf{R}$, *then we have that*:

(i) *If* $\sup_{x\in\mathbf{R}}\{G(x) - F(x)\} \le \varepsilon$, *then* $G^{-1}(1-\alpha_2) \ge F^{-1}(1-(\alpha_2+\varepsilon))$.
(ii) *If* $\sup_{x\in\mathbf{R}}\{F(x) - G(x)\} \le \varepsilon$, *then* $G^{-1}(\alpha_1) \le F^{-1}(\alpha_1+\varepsilon)$.

Furthermore, *if* $X \sim F$, *it follows that*:

(iii) *If* $\sup_{x\in\mathbf{R}}\{G(x) - F(x)\} \le \varepsilon$, *then* $P\{X \le G^{-1}(1-\alpha_2)\} \ge 1-(\alpha_2+\varepsilon)$.
(iv) *If* $\sup_{x\in\mathbf{R}}\{F(x) - G(x)\} \le \varepsilon$, *then* $P\{X \ge G^{-1}(\alpha_1)\} \ge 1-(\alpha_1+\varepsilon)$.
(v) *If* $\sup_{x\in\mathbf{R}}|G(x) - F(x)| \le \frac{\varepsilon}{2}$, *then* $P\{G^{-1}(\alpha_1) \le X \le G^{-1}(1-\alpha_2)\} \ge 1-(\alpha_1+\alpha_2+\varepsilon)$.

If $\hat{G}$ *is a random distribution function on* $\mathbf{R}$, *then we have further that*:

(vi) *If* $P\{\sup_{x\in\mathbf{R}}\{\hat{G}(x) - F(x)\} \le \varepsilon\} \ge 1-\delta$, *then* $P\{X \le \hat{G}^{-1}(1-\alpha_2)\} \ge 1-(\alpha_2+\varepsilon+\delta)$.
(vii) *If* $P\{\sup_{x\in\mathbf{R}}\{F(x) - \hat{G}(x)\} \le \varepsilon\} \ge 1-\delta$, *then* $P\{X \ge \hat{G}^{-1}(\alpha_1)\} \ge 1-(\alpha_1+\varepsilon+\delta)$.
(viii) *If* $P\{\sup_{x\in\mathbf{R}}|\hat{G}(x) - F(x)| \le \frac{\varepsilon}{2}\} \ge 1-\delta$, *then* $P\{\hat{G}^{-1}(\alpha_1) \le X \le \hat{G}^{-1}(1-\alpha_2)\} \ge 1-(\alpha_1+\alpha_2+\varepsilon+\delta)$.

PROOF. To see (i), first note that $\sup_{x\in\mathbf{R}}\{G(x) - F(x)\} \le \varepsilon$ implies that $G(x) - \varepsilon \le F(x)$ for all $x \in \mathbf{R}$. Thus, $\{x \in \mathbf{R}: G(x) \ge 1-\alpha_2\} = \{x \in \mathbf{R}: G(x) - \varepsilon \ge 1-\alpha_2-\varepsilon\} \subseteq \{x \in \mathbf{R}: F(x) \ge 1-\alpha_2-\varepsilon\}$, from which it follows that $F^{-1}(1-(\alpha_2+\varepsilon)) = \inf\{x \in \mathbf{R}: F(x) \ge 1-\alpha_2-\varepsilon\} \le \inf\{x \in \mathbf{R}: G(x) \ge 1-\alpha_2\} = G^{-1}(1-\alpha_2)$. Similarly, to prove (ii), first note that $\sup_{x\in\mathbf{R}}\{F(x) - G(x)\} \le$

ε implies that $F(x) - \varepsilon \leq G(x)$ for all $x \in \mathbf{R}$, so $\{x \in \mathbf{R} : F(x) \geq \alpha_1 + \varepsilon\} = \{x \in \mathbf{R} : F(x) - \varepsilon \geq \alpha_1\} \subseteq \{x \in \mathbf{R} : G(x) \geq \alpha_1\}$. Therefore, $G^{-1}(\alpha_1) = \inf\{x \in \mathbf{R} : G(x) \geq \alpha_1\} \leq \inf\{x \in \mathbf{R} : F(x) \geq \alpha_1 + \varepsilon\} = F^{-1}(\alpha_1 + \varepsilon)$. To prove (iii), note that because $\sup_{x \in \mathbf{R}}\{G(x) - F(x)\} \leq \varepsilon$, it follows from (i) that $\{X \leq G^{-1}(1 - \alpha_2)\} \supseteq \{X \leq F^{-1}(1 - (\alpha_2 + \varepsilon))\}$. Hence, $P\{X \leq G^{-1}(1 - \alpha_2)\} \geq P\{X \leq F^{-1}(1 - (\alpha_2 + \varepsilon))\} \geq 1 - (\alpha_2 + \varepsilon)$. Using the same reasoning, (iv) follows from (ii) and the assumption that $\sup_{x \in \mathbf{R}}\{F(x) - G(x)\} \leq \varepsilon$. To see (v), note that

$$\begin{aligned} P\{G^{-1}(\alpha_1) \leq X \leq G^{-1}(1 - \alpha_2)\} &\geq 1 - P\{X < G^{-1}(\alpha_1)\} \\ &\qquad - P\{X > G^{-1}(1 - \alpha_2)\} \\ &\geq 1 - (\alpha_1 + \alpha_2 + \varepsilon), \end{aligned}$$

where the first inequality follows from the Bonferroni inequality, and the second inequality follows from (iii) and (iv). To prove (vi), note that

$$\begin{aligned} &P\{X \leq \hat{G}^{-1}(1 - \alpha_2)\} \\ &\quad \geq P\Big\{X \leq \hat{G}^{-1}(1 - \alpha_2) \cap \sup_{x \in \mathbf{R}}\{\hat{G}(x) - F(x)\} \leq \varepsilon\Big\} \\ &\quad \geq P\Big\{X \leq F^{-1}\big(1 - (\alpha_2 + \varepsilon)\big) \cap \sup_{x \in \mathbf{R}}\{\hat{G}(x) - F(x)\} \leq \varepsilon\Big\} \\ &\quad \geq P\{X \leq F^{-1}\big(1 - (\alpha_2 + \varepsilon)\big)\} - P\Big\{\sup_{x \in \mathbf{R}}\{\hat{G}(x) - F(x)\} > \varepsilon\Big\} \\ &\quad = 1 - \alpha_2 - \varepsilon - \delta, \end{aligned}$$

where the second inequality follows from (i). A similar argument using (ii) establishes (vii). Finally, (viii) follows from (vi) and (vii) by an argument analogous to the one used to establish (v). □

LEMMA A.2. *Let $X^{(n)} = (X_1, \ldots, X_n)$ be an i.i.d. sequence of random variables with distribution P. Denote by $J_n(x, P)$ the distribution of a real-valued root $R_n = R_n(X^{(n)}, P)$ under P. Let $N_n = \binom{n}{b}$, $k_n = \lfloor \frac{n}{b} \rfloor$ and define $L_n(x, P)$ according to* (4). *Then, for any $\varepsilon > 0$, we have that*

$$P\Big\{\sup_{x \in \mathbf{R}}\big|L_n(x, P) - J_b(x, P)\big| > \varepsilon\Big\} \leq \frac{1}{\varepsilon}\sqrt{\frac{2\pi}{k_n}}. \tag{45}$$

PROOF. Let $\varepsilon > 0$ be given and define $S_n(x, P; X_1, \ldots, X_n)$ by

$$\frac{1}{k_n} \sum_{1 \leq i \leq k_n} I\big\{R_b\big((X_{b(i-1)+1}, \ldots, X_{bi}), P\big) \leq x\big\} - J_b(x, P).$$

Denote by $\mathcal{S}_n$ the symmetric group with n elements. Note that using this notation, we may rewrite $L_n(x, P) - J_b(x, P)$ as

$$Z_n(x, P; X_1, \ldots, X_n) = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} S_n(x, P; X_{\pi(1)}, \ldots, X_{\pi(n)}).$$

Note further that

$$\sup_{x \in \mathbf{R}} \bigl|Z_n(x, P; X_1, \ldots, X_n)\bigr| \le \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \sup_{x \in \mathbf{R}} \bigl|S_n(x, P; X_{\pi(1)}, \ldots, X_{\pi(n)})\bigr|,$$

which is a sum of $n!$ identically distributed random variables. Let $\varepsilon > 0$ be given. It follows that $P\{\sup_{x \in \mathbf{R}} |Z_n(x, P; X_1, \ldots, X_n)| > \varepsilon\}$ is bounded above by

$$P\biggl\{\frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \sup_{x \in \mathbf{R}} \bigl|S_n(x, P; X_{\pi(1)}, \ldots, X_{\pi(n)})\bigr| > \varepsilon\biggr\}. \tag{46}$$

Using Markov's inequality, (46) can be bounded by

$$\begin{aligned} &\frac{1}{\varepsilon} E_P\Bigl[\sup_{x \in \mathbf{R}} \bigl|S_n(x, P; X_1, \ldots, X_n)\bigr|\Bigr] \\ &\quad = \frac{1}{\varepsilon} \int_0^1 P\Bigl\{\sup_{x \in \mathbf{R}} \bigl|S_n(x, P; X_1, \ldots, X_n)\bigr| > u\Bigr\}\, du. \end{aligned} \tag{47}$$

We may use the Dvoretsky–Kiefer–Wolfowitz inequality to bound the right-hand side of (47) by

$$\frac{1}{\varepsilon} \int_0^1 2 \exp\{-2k_n u^2\}\, du = \frac{2}{\varepsilon} \sqrt{\frac{2\pi}{k_n}} \biggl[\Phi(2\sqrt{k_n}) - \frac{1}{2}\biggr] < \frac{1}{\varepsilon} \sqrt{\frac{2\pi}{k_n}},$$

which establishes (45). □

LEMMA A.3. *Let $X^{(n)} = (X_1, \ldots, X_n)$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$. Denote by $J_n(x, P)$ the distribution of a real-valued root $R_n = R_n(X^{(n)}, P)$ under P. Let $k_n = \lfloor \frac{n}{b} \rfloor$ and define $L_n(x, P)$ according to* (4). *Let*

$$\delta_{1,n}(\varepsilon, \gamma, P) = \frac{1}{\gamma\varepsilon} \sqrt{\frac{2\pi}{k_n}} + I\Bigl\{\sup_{x \in \mathbf{R}} \{J_b(x, P) - J_n(x, P)\} > (1 - \gamma)\varepsilon\Bigr\},$$

$$\delta_{2,n}(\varepsilon, \gamma, P) = \frac{1}{\gamma\varepsilon} \sqrt{\frac{2\pi}{k_n}} + I\Bigl\{\sup_{x \in \mathbf{R}} \{J_n(x, P) - J_b(x, P)\} > (1 - \gamma)\varepsilon\Bigr\},$$

$$\delta_{3,n}(\varepsilon, \gamma, P) = \frac{1}{\gamma\varepsilon} \sqrt{\frac{2\pi}{k_n}} + I\Bigl\{\sup_{x \in \mathbf{R}} \bigl|J_b(x, P) - J_n(x, P)\bigr| > (1 - \gamma)\varepsilon\Bigr\}.$$

Then, for any $\varepsilon > 0$ and $\gamma \in (0, 1)$, we have that:

(i) $P\{R_n \le L_n^{-1}(1-\alpha_2, P)\} \ge 1-(\alpha_2+\varepsilon+\delta_{1,n}(\varepsilon,\gamma,P))$;
(ii) $P\{R_n \ge L_n^{-1}(\alpha, P)\} \ge 1-(\alpha_1+\varepsilon+\delta_{2,n}(\varepsilon,\gamma,P))$;
(iii) $P\{L_n^{-1}(\alpha_1, P) \le R_n \le L_n^{-1}(1-\alpha_2, P)\} \ge 1-(\alpha_1+\alpha_2+\varepsilon+\delta_{3,n}(\varepsilon, \gamma, P))$.

PROOF. Let $\varepsilon > 0$ and $\gamma \in (0,1)$ be given. Note that

$$\begin{aligned}
&P\Big\{\sup_{x\in\mathbf{R}}\{L_n(x,P)-J_n(x,P)\} > \varepsilon\Big\} \\
&\quad \le P\Big\{\sup_{x\in\mathbf{R}}\{L_n(x,P)-J_b(x,P)\} + \sup_{x\in\mathbf{R}}\{J_b(x,P)-J_n(x,P)\} > \varepsilon\Big\} \\
&\quad \le P\Big\{\sup_{x\in\mathbf{R}}\{L_n(x,P)-J_b(x,P)\} > \gamma\varepsilon\Big\} \\
&\qquad + I\Big\{\sup_{x\in\mathbf{R}}\{J_b(x,P)-J_n(x,P)\} > (1-\gamma)\varepsilon\Big\} \\
&\quad \le \frac{1}{\gamma\varepsilon}\sqrt{\frac{2\pi}{k_n}} + I\Big\{\sup_{x\in\mathbf{R}}\{J_b(x,P)-J_n(x,P)\} > (1-\gamma)\varepsilon\Big\},
\end{aligned}$$

where the final inequality follows from Lemma A.2. Assertion (i) thus follows from the definition of $\delta_{1,n}(\varepsilon,\gamma,P)$ and part (vi) of Lemma A.1. Assertions (ii) and (iii) are established similarly. □

PROOF OF THEOREM 2.1. To prove (i), note that by part (i) of Lemma A.3, we have for any $\varepsilon > 0$ and $\gamma \in (0,1)$ that

$$\sup_{P\in\mathbf{P}} P\{R_n \le L_n^{-1}(1-\alpha_2, P)\} \ge 1-\Big(\alpha_2+\varepsilon+\inf_{P\in\mathbf{P}}\delta_{1,n}(\varepsilon,\gamma,P)\Big),$$

where

$$\delta_{1,n}(\varepsilon,\gamma,P) = \frac{1}{\gamma\varepsilon}\sqrt{\frac{2\pi}{k_n}} + I\Big\{\sup_{x\in\mathbf{R}}\{J_b(x,P)-J_n(x,P)\} > (1-\gamma)\varepsilon\Big\}.$$

By the assumption on $\sup_{P\in\mathbf{P}}\sup_{x\in\mathbf{R}}\{J_b(x,P)-J_n(x,P)\}$, we have that $\inf_{P\in\mathbf{P}}\delta_{1,n}(\varepsilon,\gamma,P) \to 0$ for every $\varepsilon > 0$. Thus, there exists a sequence $\varepsilon_n > 0$ tending to 0 so that $\inf_{P\in\mathbf{P}}\delta_{1,n}(\varepsilon_n,\gamma,P) \to 0$. The desired claim now follows from applying part (i) of Lemma A.3 to this sequence. Assertions (ii) and (iii) follow in exactly the same way. □

A.2. Proof of Theorem 2.4. We prove only (i). Similar arguments can be used to establish (ii) and (iii). Let $\alpha_1 = 0$, $0 \le \alpha_2 < 1$ and $\eta > 0$ be given. Choose $\delta > 0$ so that

$$\sup_{x\in\mathbf{R}}\{J_n(x,P') - J_n(x,P)\} < \frac{\eta}{2},$$

whenever $\rho(P', P) < \delta$ for $P' \in \mathbf{P}'$ and $P \in \mathbf{P}$. For n sufficiently large, we have that

$$\sup_{P \in \mathbf{P}} P\{\rho(\hat{P}_n, P) > \delta\} < \frac{\eta}{4} \quad \text{and} \quad \sup_{P \in \mathbf{P}} P\{\hat{P}_n \notin \mathbf{P}'\} < \frac{\eta}{4}.$$

For such n, we therefore have that

$$\begin{aligned} 1 - \frac{\eta}{2} &\leq \inf_{P \in \mathbf{P}} P\{\rho(\hat{P}_n, P) \leq \delta \cap \hat{P}_n \in \mathbf{P}'\} \\ &\leq \inf_{P \in \mathbf{P}} P\left\{\sup_{x \in \mathbf{R}}\{J_n(x, \hat{P}_n) - J_n(x, P)\} \leq \frac{\eta}{2}\right\}. \end{aligned}$$

It follows from part (vi) of Lemma A.1 that for such n

$$\inf_{P \in \mathbf{P}} P\{R_n \leq J_n^{-1}(1 - \alpha_2, \hat{P}_n)\} \geq 1 - (\alpha_2 + \eta).$$

Since the choice of η was arbitrary, the desired result follows.

SUPPLEMENTARY MATERIAL

Supplement to "On the uniform asymptotic validity of subsampling and the bootstrap" (DOI: 10.1214/12-AOS1051SUPP; .pdf). The supplement provides additional details and proofs for many of the results in the authors' paper.

REFERENCES

ANDREWS, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68** 399–405. MR1748009

ANDREWS, D. W. K. and BARWICK, P. J. (2012). Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica* **80** 2805–2826.

ANDREWS, D. W. K. and GUGGENBERGER, P. (2009). Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econometric Theory* **25** 669–709. MR2507528

ANDREWS, D. W. K. and GUGGENBERGER, P. (2010). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory* **26** 426–468. MR2600570

ANDREWS, D. W. K. and SOARES, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78** 119–157. MR2642858

BAHADUR, R. R. and SAVAGE, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27** 1115–1122. MR0084241

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York. MR0436272

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

BUGNI, F. A. (2010). Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica* **78** 735–753. MR2656646

CANAY, I. A. (2010). EL inference for partially identified models: Large deviations optimality and bootstrap validity. *J. Econometrics* **156** 408–425. MR2609942

KABAILA, P. (1995). The effect of model selection on confidence regions and prediction regions. *Econometric Theory* **11** 537–549. MR1349934

LEEB, H. and PÖTSCHER, B. M. (2006a). Can one estimate the conditional distribution of post-model-selection estimators? *Ann. Statist.* **34** 2554–2591. MR2291510

LEEB, H. and PÖTSCHER, B. M. (2006b). Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk-bound results. *Econometric Theory* **22** 69–97. MR2212693

MIKUSHEVA, A. (2007). Uniform inference in autoregressive models. *Econometrica* **75** 1411–1452. MR2347350

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York. MR1707286

PÖTSCHER, B. M. (2002). Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* **70** 1035–1065. MR1910411

PÖTSCHER, B. M. (2009). Confidence sets based on sparse estimators are necessarily large. *Sankhyā* **71** 1–18. MR2579644

ROMANO, J. P. (1989). Do bootstrap confidence procedures behave well uniformly in P? *Canad. J. Statist.* **17** 75–80. MR1014092

ROMANO, J. P. and SHAIKH, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *J. Statist. Plann. Inference* **138** 2786–2807. MR2422399

ROMANO, J. P. and SHAIKH, A. M. (2010). Inference for the identified set in partially identified econometric models. *Econometrica* **78** 169–211. MR2642860

ROMANO, J. P. and SHAIKH, A. M. (2012). Supplement to "On the uniform asymptotic validity of subsampling and the bootstrap." DOI:10.1214/12-AOS1051SUPP.

ROMANO, J. P., SHAIKH, A. M. and WOLF, M. (2012). A simple two-step approach to testing moment inequalities with an application to inference in partially identified models. Working paper.

DEPARTMENTS OF ECONOMICS
AND STATISTICS
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CALIFORNIA 94305-4065
USA
E-MAIL: romano@stanford.edu

DEPARTMENT OF ECONOMICS
UNIVERSITY OF CHICAGO
1126 E. 59TH STREET
CHICAGO, ILLINOIS 60637
USA
E-MAIL: amshaikh@uchicago.edu