ECONOMETRICA

INFERENCE FOR THE IDENTIFIED SET IN PARTIALLY IDENTIFIED ECONOMETRIC MODELS

JOSEPH P. ROMANO
Stanford University, Stanford, CA 94305-4065, U.S.A.

AZEEM M. SHAIKH
University of Chicago, Chicago, IL 60637, U.S.A.

INFERENCE FOR THE IDENTIFIED SET IN PARTIALLY IDENTIFIED ECONOMETRIC MODELS

By Joseph P. Romano and Azeem M. Shaikh[1]

This paper provides computationally intensive, yet feasible methods for inference in a very general class of partially identified econometric models. Let P denote the distribution of the observed data. The class of models we consider is defined by a population objective function $Q(\theta, P)$ for $\theta \in \Theta$. The point of departure from the classical extremum estimation framework is that it is not assumed that $Q(\theta, P)$ has a unique minimizer in the parameter space Θ. The goal may be either to draw inferences about some unknown point in the set of minimizers of the population objective function or to draw inferences about the set of minimizers itself. In this paper, the object of interest is $\Theta_0(P) = \arg\min_{\theta \in \Theta} Q(\theta, P)$, and so we seek random sets that contain this set with at least some prespecified probability asymptotically. We also consider situations where the object of interest is the image of $\Theta_0(P)$ under a known function. Random sets that satisfy the desired coverage property are constructed under weak assumptions. Conditions are provided under which the confidence regions are asymptotically valid not only pointwise in P, but also uniformly in P. We illustrate the use of our methods with an empirical study of the impact of top-coding outcomes on inferences about the parameters of a linear regression. Finally, a modest simulation study sheds some light on the finite-sample behavior of our procedure.

KEYWORDS: Partially identified model, incomplete model, identified set, identifiable parameter, subsampling, uniform coverage, confidence region, moment inequalities.

1. INTRODUCTION

A PARTIALLY IDENTIFIED MODEL is any model in which the parameter of interest is not uniquely defined by the distribution of the observed data. This paper provides computationally intensive yet feasible methods for inference for one large class of such models. Let P denote the distribution of the observed data. The class of models we consider is defined by a population objective function $Q(\theta, P)$ for $\theta \in \Theta$. The point of departure from the classical extremum estimation framework is that it is not assumed that $Q(\theta, P)$ has a unique minimizer in the parameter space Θ. The goal may be either to draw inferences about some unknown point in the set of minimizers of the population objective function or to draw inferences about the set of minimizers itself. In this paper we consider the second of these two goals. The object of interest is

$$(1) \qquad \Theta_0(P) = \arg\min_{\theta \in \Theta} Q(\theta, P).$$

We henceforth refer to $\Theta_0(P)$ as the *identified set*. In this instance, given independent and identically distributed (i.i.d.) data $X_i, i = 1, \ldots, n$, generated

from P, we seek random sets $\mathcal{C}_n = \mathcal{C}_n(X_1, \ldots, X_n)$ that contain the identified set with at least some prespecified probability asymptotically. That is, we require

$$(2) \qquad \liminf_{n \to \infty} P\{\Theta_0(P) \subseteq \mathcal{C}_n\} \geq 1 - \alpha.$$

We refer to such sets as *confidence regions for the identified set that are pointwise consistent in level*. This terminology reflects the fact that the confidence regions are valid only for a *fixed* probability distribution P and helps distinguish this coverage requirement from others discussed later in which we will demand that the confidence regions are valid uniformly in P. We show that the problem of constructing $\mathcal{C}_n$ that satisfy (2) is equivalent to a multiple hypothesis testing problem in which one wants to test the family of null hypotheses $H_\theta : \theta \in \Theta_0(P)$ indexed by $\theta \in \Theta$ while controlling the familywise error rate, the probability of even one false rejection under P. Using this duality, we go on to construct $\mathcal{C}_n$ that satisfy (2) under weak assumptions on P.

In the first goal, the object of interest is some unknown point $\theta \in \Theta_0(P)$. We refer to any $\theta \in \Theta_0(P)$ as an *identifiable parameter*. In this case, given i.i.d. data X_i, $i = 1, \ldots, n$, generated from P, we seek random sets $\mathcal{C}_n = \mathcal{C}_n(X_1, \ldots, X_n)$ that contain each identifiable parameter with at least some prespecified probability asymptotically. The problem of constructing such sets is treated in a companion paper (Romano and Shaikh (2008)).

Our results on confidence regions for the identified set build upon the earlier work of Chernozhukov, Hong, and Tamer (2007), who were the first to consider inference for the same class of partially identified models. An important feature of our procedure for constructing confidence regions for the identified set is that it avoids the need for an initial estimate of $\Theta_0(P)$. In general, our procedure is first-order asymptotically equivalent with the procedure proposed by Chernozhukov, Hong, and Tamer (2007). On the other hand, when the set of minimizers of $\hat{Q}_n(\theta)$ does not provide a consistent estimate of $\Theta_0(P)$, our results provide a justification for iterating their procedure until a stopping criterion is met to produce confidence regions that are typically strictly smaller while still maintaining the coverage requirement.

In this paper, we also wish to construct confidence regions whose coverage probability is close to the nominal level not just for a fixed probability distribution P, but rather uniformly over all P in some large class of distributions $\mathbf{P}$. Confidence regions that fail to satisfy this requirement have the feature that for every sample size n, however large, there is some probability distribution $P \in \mathbf{P}$ for which the coverage probability of the confidence region under P is not close to the prescribed level. Researchers may, therefore, feel that inferences made on the basis of asymptotic approximations are more reliable if the confidence regions exhibit good uniform behavior. Of course, such a requirement will typically require restrictions on P beyond those required for pointwise consistency in level. Bahadur and Savage (1956), for example, showed that if $\mathbf{P}$ is suitably large, then there exists no confidence interval for the mean with finite length

and good uniform behavior. Romano (2004) extended this nonexistence result to a number of other problems. We provide restrictions on **P** under which the confidence regions in this paper have good uniform behavior. Concretely, we provide conditions under which $\mathcal{C}_n$ satisfies

$$(3) \qquad \liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P\{\Theta_0(P) \subseteq \mathcal{C}_n\} \geq 1 - \alpha.$$

By analogy with our earlier terminology, sets that satisfy (3) are referred to as *confidence regions for the identified set that are uniformly consistent in level*. Note that if the identified set $\Theta_0(P)$ consists of a single point $\theta_0(P)$, then this definition reduces to the usual definition of confidence regions that are uniformly consistent in level; that is,

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P\{\theta_0(P) \in \mathcal{C}_n\} \geq 1 - \alpha.$$

Imbens and Manski (2004) analyzed the special case of the above class of partially identified models in which the identified set is an interval whose upper and lower endpoints are means or at least behave like means asymptotically. For this special case, they constructed confidence regions that contain each identifiable parameter with at least some prespecified probability asymptotically and are valid uniformly in P. Romano and Shaikh (2008) constructed confidence regions with this same coverage property for the more general class of models considered here. To the best of our knowledge, this paper is the first to consider confidence regions for the identified set that are valid uniformly in P.

We have so far assumed that the object of interest is the identified set, $\Theta_0(P)$, itself. More generally, the object of interest may be the image of the identified set under a known function. A typical example of such a function is the projection of $\mathbf{R}^k$ onto one of the axes. We extend the above definitions of confidence regions to this setting as follows. Consider a function $f : \Theta \to \Lambda$. Denote by $\Lambda_0(P)$ the image of $\Theta_0(P)$ under f; that is,

$$(4) \qquad \Lambda_0(P) = \{f(\theta) : \theta \in \Theta_0(P)\}.$$

We refer to a set $\mathcal{C}_n^f$ as a *confidence region for a function of the identified set that is pointwise consistent in level* if it satisfies

$$(5) \qquad \liminf_{n\to\infty} P\{\Lambda_0(P) \in \mathcal{C}_n^f\} \geq 1 - \alpha.$$

As before, we may also demand uniformly good behavior over a class of probability distributions **P** by requiring that $\mathcal{C}_n^f$ satisfy

$$(6) \qquad \liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P\{\Lambda_0(P) \in \mathcal{C}_n^f\} \geq 1 - \alpha.$$

By analogy with our earlier terminology, sets that satisfy (6) are referred to as *confidence regions for a function of the identified set that are uniformly consistent*

in level. We adapt our constructions of confidence regions for the identified set to provide constructions of confidence sets that satisfy these alternative coverage requirements.

The remainder of the paper is organized as follows. In Section 2, we consider the problem of constructing confidence regions that satisfy the coverage properties (2) and (3). The construction exploits a useful equivalence between the construction of confidence regions for the identified set and a suitable multiple hypothesis testing problem. We then extend this methodology to construct confidence regions that satisfy (5) and (6). We provide an illustration of our methods in Section 3. In Section 4, we shed some light on the finite-sample behavior of our methodology via a small simulation study. Data and programs are provided as Supplemental Material (Romano and Shaikh (2010)).

2. CONFIDENCE REGIONS FOR THE IDENTIFIED SET

In this section, we consider the problem of constructing confidence regions for the identified set. We begin by treating the construction of sets that satisfy (2) before turning our attention to the problem of constructing sets that satisfy (3).

2.1. *Pointwise Consistency in Level*

2.1.1. *Equivalence With a Multiple Testing Problem*

We will first show that the problem of constructing confidence regions that satisfy (2) is equivalent to a certain multiple hypothesis testing problem. The problem is to test the family of hypotheses

$$(7) \qquad H_\theta : Q(\theta, P) = 0 \quad \text{for} \quad \theta \in \Theta$$

in a way that asymptotically controls the familywise error rate (FWER_P), the probability of one or more false rejections under P, at level α. Formally,

$$(8) \qquad \mathrm{FWER}_P = P\{\text{reject at least 1 null hypothesis } H_\theta \text{ s.t. } Q(\theta, P) = 0\},$$

and by asymptotic control of the FWER_P at level α, we mean the requirement that

$$(9) \qquad \limsup_{n \to \infty} \mathrm{FWER}_P \le \alpha.$$

The following lemma establishes the equivalence between these two problems.

LEMMA 2.1: *Let P denote the true distribution of the data. Given any procedure for testing the family of null hypotheses* (7) *which yields a decision for each of the*

null hypotheses, the set of θ values for which the corresponding null hypothesis H_θ is accepted, C_n, satisfies

$$P\{\Theta_0(P) \subseteq C_n\} = 1 - \text{FWER}_P,$$

where $\Theta_0(P)$ is defined by (1). *Conversely, given any random set C_n, the procedure for testing the family of hypotheses* (7) *in which a null hypothesis H_θ is accepted if and only if $\theta \in C_n$ satisfies*

$$\text{FWER}_P = 1 - P\{\Theta_0(P) \subseteq C_n\}.$$

PROOF: To establish the first conclusion, note that by the definition of $\Theta_0(P)$ we have

$$P\{\Theta_0(P) \subseteq C_n\} = P\{\text{reject no null hypothesis } H_\theta \text{ s.t. } Q(\theta, P) = 0\}$$
$$= 1 - \text{FWER}_P.$$

The second conclusion follows from the same reasoning. *Q.E.D.*

It follows from Lemma 2.1 that given any procedure for testing the family of null hypotheses (7) that satisfy (9), the set of θ values corresponding to the set of accepted hypotheses, C_n, satisfies (2). We thus turn to the problem of constructing tests of (7) that satisfy (9).

2.1.2. *Single-Step Control of the Familywise Error Rate*

First, we briefly discuss a single-step approach to asymptotic control of the FWER_P at level α, since it serves as a building block for the more powerful stepdown procedures that we will develop in the next section. As before, we will require a test statistic for each null hypothesis H_θ such that large values of the test statistic provide evidence against the null hypothesis. The statistic $a_n \hat{Q}_n(\theta)$ for some sequence $a_n \to \infty$ will be used for this purpose. We assume that the sequence a_n is known and that it does not depend on θ, but both of these requirements can be relaxed using ideas in Chapter 8 of Politis, Romano, and Wolf (1999).

For $K \subseteq \Theta$, let $c_n(K, 1 - \alpha, P)$ denote the smallest $1 - \alpha$ quantile of the distribution of

$$(10) \qquad \sup_{\theta \in K} a_n \hat{Q}_n(\theta)$$

under P; that is,

$$c_n(K, 1 - \alpha, P) = \inf\left\{ x : P\left\{ \sup_{\theta \in K} a_n \hat{Q}_n(\theta) \le x \right\} \ge 1 - \alpha \right\}.$$

Consider the idealized test in which a null hypothesis H_θ is rejected if and only if $a_n \hat{Q}_n(\theta) > c_n(\Theta_0(P), 1 - \alpha, P)$. This is a single-step method in the sense that each $a_n \hat{Q}_n(\theta)$ is compared with a common value so as to determine its significance. Clearly, such a test satisfies $\text{FWER}_P \leq \alpha$. To see this, note that

$$\text{FWER}_P = P\{a_n \hat{Q}_n(\theta) > c_n(\Theta_0(P), 1 - \alpha, P) \text{ for some } \theta \in \Theta_0(P)\}$$
$$= 1 - P\left\{\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \leq c_n(\Theta_0(P), 1 - \alpha, P)\right\} \leq \alpha.$$

But this test is infeasible, as the critical value depends on the unknown P. A crude solution to this difficulty is available if estimators $\hat{c}_n(K, 1 - \alpha)$ of $c_n(K, 1 - \alpha, P)$ are available that satisfy two properties. First, we require that $\hat{c}_n(K, 1 - \alpha)$ be a "good" estimator when $K = \Theta_0(P)$ in the sense that

$$(11) \qquad \limsup_{n \to \infty} P\left\{\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) > \hat{c}_n(\Theta_0(P), 1 - \alpha)\right\} \leq \alpha.$$

Second, we require that the estimators be monotone in the sense that

$$(12) \qquad \hat{c}_n(K, 1 - \alpha) \geq \hat{c}_n(\Theta_0(P), 1 - \alpha) \quad \text{for any } K \supseteq \Theta_0(P).$$

Since $\Theta_0(P) \subseteq \Theta$, it follows that under these assumptions we have that $\hat{c}_n(\Theta, 1 - \alpha)$ asymptotically provides a conservative estimator of $c_n(\Theta_0(P), 1 - \alpha, P)$. Hence, the single-step method in which each statistic $a_n \hat{Q}_n(\theta)$ is compared with the common cutoff $\hat{c}_n(\Theta, 1 - \alpha)$ asymptotically controls the FWER_P at level α provided that these two assumptions are satisfied. We will refrain from stating this result more formally because it will follow from the analysis of the more powerful step-down method in the next section. We will also provide an explicit construction of such estimators in the subsequent section.

2.1.3. *Step-Down Control of the Familywise Error Rate*

Step-down methods begin by first applying a single-step method, but then additional hypotheses may be rejected after this first stage by proceeding in a stepwise fashion, which we now describe. In the first stage, test the entire family of hypotheses using a single-step procedure; that is, reject all null hypotheses whose corresponding test statistic is too large, where large is determined by some common critical value as described above. If no hypotheses are rejected in this first stage, then stop; otherwise, test the family of hypotheses not rejected in the first stage using a single-step procedure. If no further hypotheses are rejected in this second stage, then stop; otherwise, test the family of hypotheses not rejected in the first and second stages using a single-step procedure. Repeat this process until no further hypotheses are rejected. We now formally define this procedure, which can be viewed as a generalization

of Romano and Wolf (2005), who only considered a finite number of hypotheses.

ALGORITHM 2.1:
1. Let $S_1 = \Theta$. If $\sup_{\theta \in S_1} a_n \hat{Q}_n(\theta) \leq \hat{c}_n(S_1, 1 - \alpha)$, then accept all hypotheses and stop; otherwise, set $S_2 = \{\theta \in \Theta : a_n \hat{Q}_n(\theta) \leq \hat{c}_n(S_1, 1 - \alpha)\}$ and continue.
$\vdots$
j. If $\sup_{\theta \in S_j} a_n \hat{Q}_n(\theta) \leq \hat{c}_n(S_j, 1 - \alpha)$, then accept all hypotheses H_θ with $\theta \in S_j$ and stop; otherwise, set $S_{j+1} = \{\theta \in \Theta : a_n \hat{Q}_n(\theta) \leq \hat{c}_n(S_j, 1 - \alpha)\}$ and continue.
$\vdots$

We now prove that this algorithm provides asymptotic control of the FWER_P under the monotonicity assumption (11) and (12).

THEOREM 2.1: *Let P denote the true distribution generating the data. Consider Algorithm* 2.1 *with critical values that satisfy* (12). *Then*

$$(13) \qquad \text{FWER}_P \leq P\Big\{ \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) > \hat{c}_n(\Theta_0(P), 1 - \alpha) \Big\}.$$

Hence, if the critical values also satisfy (11), *then*

$$\limsup_{n \to \infty} \text{FWER}_P \leq \alpha.$$

PROOF: To establish (13), denote by $\hat{j}$ the smallest random index for which there is a false rejection; that is, there exists $\theta' \in \Theta_0(P)$ such that $a_n \hat{Q}_n(\theta') > \hat{c}_n(S_{\hat{j}}, 1 - \alpha)$. By definition of $\hat{j}$, we must have that $\Theta_0(P) \subseteq S_{\hat{j}}$. Thus, by (12) we have that $\hat{c}_n(S_{\hat{j}}, 1 - \alpha) \geq \hat{c}_n(\Theta_0(P), 1 - \alpha)$. Hence, it must be the case that

$$\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \geq a_n \hat{Q}_n(\theta') > \hat{c}_n(\Theta_0(P), 1 - \alpha).$$

The second conclusion follows immediately.   *Q.E.D.*

2.1.4. *A Subsampling Construction*

It follows from Theorem 2.1 that under the two restrictions (11) and (12), the set of θ values corresponding to the accepted hypotheses from Algorithm 2.1, $\mathcal{C}_n$, satisfies (2). We now provide a concrete construction of critical values that satisfy these two properties under a weak assumption on the asymptotic behavior of the test statistics $a_n \hat{Q}_n(\theta)$.

The construction will be based on subsampling. To define the critical values precisely, some further notation is required. Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$. Let $N_n = \binom{n}{b}$ and let $\hat{Q}_{n,b,i}(\theta)$ denote the statistic $\hat{Q}_n(\theta)$ evaluated at the ith subset of data of size b from the n observations. For $K \subseteq \Theta$ and $\alpha \in (0, 1)$, define

$$(14) \qquad \hat{r}_n(K, 1 - \alpha) = \inf\left\{ x : \frac{1}{N_n} \sum_{1 \le i \le N_n} I\left\{ \sup_{\theta \in K} a_b \hat{Q}_{n,b,i}(\theta) \le x \right\} \ge 1 - \alpha \right\}.$$

Note that by construction, the critical values defined by (14) satisfy the monotonicity restriction (12). We now provide conditions under which they also satisfy (11).

THEOREM 2.2: *Let X_i, $i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P and let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$. Let $J_n(\cdot, P)$ denote the distribution of $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta)$ under P. Suppose $J_n(\cdot, P)$ converges in distribution to a limit distribution $J(\cdot, P)$ and that $J(\cdot, P)$ is continuous at its smallest $1 - \alpha$ quantile. Then the following statements are true*:

(i) *Condition (11) holds when $\hat{c}_n(\Theta_0(P), 1 - \alpha)$ is given by (14) with $K = \Theta_0(P)$.*

(ii) *Algorithm 2.1 with $\hat{c}_n(K, 1 - \alpha)$ given by (14) provides asymptotic control of the* FWER$_P$ *at level α.*

(iii) *The set of θ values corresponding to accepted hypotheses from Algorithm 2.1 with $\hat{c}_n(K, 1 - \alpha)$ given by (14), $\mathcal{C}_n$, satisfies (2).*

PROOF: The first result follows from Theorem 2.1.1 of Politis, Romano, and Wolf (1999). The second follows from Theorem 2.1. The third follows from Lemma 2.1.   *Q.E.D.*

REMARK 2.1: Because $\binom{n}{b}$ may be large, it is often more practical to use the following approximation to (14). Let the sequence $B_n \to \infty$ as $n \to \infty$ and let $I_1, \ldots, I_{B_n}$ be chosen randomly with or without replacement from the numbers $1, \ldots, N_n$. Then it follows from Corollary 2.4.1 of Politis, Romano, and Wolf (1999) that one may approximate (14) by

$$\inf\left\{ x : \frac{1}{B_n} \sum_{1 \le i \le B_n} I\left\{ \sup_{\theta \in K} a_b \hat{Q}_{n,b,I_i}(\theta) \le x \right\} \ge 1 - \alpha \right\}$$

without affecting the conclusions of Theorem 2.2.

REMARK 2.2: Throughout this paper, we assume that the observations X_i, $i = 1, \ldots, n$, are an i.i.d. sequence of random variables with distribution P.

Many of the results, however, can be extended to certain time series settings. Consider, for example, the following extension of Theorem 2.2. Let

$$(15) \quad \hat{r}_n(K, 1 - \alpha)$$

$$= \inf\left\{x : \frac{1}{n - b + 1} \sum_{1 \le i \le n - b + 1} I\left\{\sup_{\theta \in K} a_b \hat{Q}_{n,b,i}(\theta) \le x\right\} \ge 1 - \alpha\right\},$$

where $i = 1, \ldots, n - b + 1$ now indexes only the subsets of data of size b whose observations are consecutive. If one assumes that the $X_i, i = 1, \ldots, n$, are observations from a distribution P for which the corresponding α-mixing sequence $\alpha_X(m) \to 0$ as $m \to \infty$, but otherwise maintains the assumptions of Theorem 2.2, then it follows from Theorem 3.2.1 of Politis, Romano, and Wolf (1999) that the conclusions of the theorem continue to hold.

We now consider two important examples and use Theorem 2.2 to provide conditions under which Algorithm 2.1 with $\hat{c}_n(K, 1 - \alpha)$ given by (14) asymptotically controls the FWER$_P$ and thus the set of θ values corresponding to the accepted hypotheses, $\mathcal{C}_n$, satisfies (2).

EXAMPLE 2.1—Moment Inequalities: Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P on $\mathbf{R}^k$. For $j = 1, \ldots, m$, let $g_j(x, \theta)$ be a real-valued function on $\mathbf{R}^k \times \mathbf{R}^l$. The identified set is assumed to be $\Theta_0(P) = \{\theta \in \mathbf{R}^l : E_P[g_j(X_i, \theta)] \le 0 \; \forall j \text{ s.t. } 1 \le j \le m\}$. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = \sum_{1 \le j \le m} \left(E_P[g_j(X_i, \theta)]\right)_+^2,$$

where $(x)_+ = \max\{x, 0\}$. The sample analog of $Q(\theta, P)$ is given by

$$\hat{Q}_n(\theta) = \sum_{1 \le j \le m} \left(\frac{1}{n} \sum_{1 \le i \le n} g_j(X_i, \theta)\right)_+^2.$$

Let $a_n = n$ and suppose P is such that (i)

$$\{g_j(\cdot, \theta) : 1 \le j \le m, \theta \in \Theta_0(P)\} \quad \text{is } P\text{-Donsker},$$

(ii) for every $K \subseteq \{1, \ldots, m\}$ and sequence $\theta_n \in \Theta_0(P)$ such that $E_P[g_j(X_i, \theta_n)] \to 0$ for all $j \in K$, there exists a subsequence n_k and a $\theta \in \Theta_0(P)$ such that $E_P[g_j(X_i, \theta)] = 0$ for all $j \in K$ and $\rho_{P,j}(\theta_{n_k}, \theta) \to 0$ for all $j \in K$, where $\rho_{P,j}^2(\theta, \theta') = E_P[(g_j(X_i, \theta) - g_j(X_i, \theta'))^2] \to 0$. To rule out degenerate situations, assume further that (iii) there exist $1 \le j^* \le m$ and $\theta^* \in \Theta_0(P)$ such that $E_P[g_{j^*}(X_i, \theta^*)] = 0$ and $\text{Var}_P[g_{j^*}(X_i, \theta^*)] > 0$. Assumption (i) is known to hold provided that the class of functions is not too large; for general results to this

end and numerous applications, see van der Vaart and Wellner (1996). The conditions of Theorem 2.2 are verified for $\alpha < \frac{1}{2}$ under these assumptions in Section A.2 of the Appendix.

EXAMPLE 2.2—Regression With Interval Outcomes: The following example allows for inference in a linear regression model in which the dependent variable is interval-censored. Let $(X_i, Y_{1,i}, Y_{2,i}, Y_i^*)$, $i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P^* on $\mathbf{R}^k \times \mathbf{R} \times \mathbf{R} \times \mathbf{R}$. The parameter of interest, θ_0, is known to satisfy $E_{P^*}[Y_i^* | X_i] = X_i' \theta_0$, but Y_i^* is unobserved, which precludes conventional estimation of θ_0. Let P denote the distribution of the observed random variables $(X_i, Y_{1,i}, Y_{2,i})$. The random variables $(Y_{1,i}, Y_{2,i})$ are known to satisfy $Y_{1,i} \leq Y_i^* \leq Y_{2,i}$ with probability 1 under P^*. Thus, $\theta_0 \in \Theta_0(P) = \{\theta \in \mathbf{R}^k : E_P[Y_{1,i} | X_i] \leq X_i' \theta \leq E_P[Y_{2,i} | X_i] \; P\text{-a.s.}\}$. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = E_P\big[(E_P[Y_{1,i} | X_i] - X_i' \theta)_+^2 + (X_i' \theta - E_P[Y_{2,i} | X_i])_+^2\big].$$

Manski and Tamer (2002) characterized the identified set in this setting and also considered the case where Y_i^* is observed, but X_i is interval-censored.

Let $a_n = n$ and suppose P is such that (i) $\text{supp}_P(X_i) = \{x_1, \ldots, x_m\}$ and (ii) the variances of Y_1 and Y_2, $\sigma_1^2(P)$ and $\sigma_2^2(P)$, exist. To rule out degenerate situations, assume further that (iii) there exist $\theta^* \in \Theta$, $\ell^* \in \{1, 2\}$, and $j^* \in \{1, \ldots, m\}$ such that $E_P[Y_{\ell^*,i} | X_i = x_{j^*}] = x_{j^*}' \theta^*$ and $\text{Var}_P[Y_{\ell^*,i} | X_i = x_{j^*}] > 0$. For $\ell \in \{1, 2\}$ and $j \in \{1, \ldots, m\}$, let $\tau_\ell(x_j, P) = E_P[Y_{\ell,i} | X_i = x_j]$ and

$$\hat{\tau}_\ell(x_j) = \frac{1}{n(x_j)} \sum_{1 \leq i \leq n : X_i = x_j} Y_{l,i},$$

where $n(x_j) = |\{1 \leq i \leq n : X_i = x_j\}|$. Let

$$\hat{Q}_n(\theta) = \sum_{1 \leq j \leq m} \frac{n(x_j)}{n} \big\{(\hat{\tau}_1(x_j) - x_j' \theta)_+^2 + (x_j' \theta - \hat{\tau}_2(x_j))_+^2\big\}.$$

We now verify the conditions of Theorem 2.2 under these assumptions for $\alpha < \frac{1}{2}$.

To this end, note that

$$a_n \hat{Q}_n(\theta) = \sum_{1 \leq j \leq m} \sum_{1 \leq \ell \leq 2} \bigg(\sqrt{\frac{n}{n(x_j)}} \frac{1}{\sqrt{n}}$$

$$\times \sum_{1 \leq i \leq n} (-1)^{\ell-1} (Y_{\ell,i} - x_j' \theta) I\{X_i = x_j\}\bigg)_+^2.$$

For $K \subseteq \{1, \ldots, m\} \times \{1, 2\}$ let

$$\Theta_0(K, P) = \left\{ \theta \in \Theta_0(P) : E_P[Y_{\ell,i}|X_i = x_j] = x_j'\theta \text{ for all } (j, \ell) \in K \right\}.$$

Hence, except for the multiplicative factors $\sqrt{n/n(x_j)}$, which are asymptotically constant anyway, the structure here is the same as the structure of Example 2.1. As a result, we may use arguments nearly identical to those given for Example 2.1 above in Section A.2 of the Appendix to show that the limiting behavior of

$$(16) \qquad \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta)$$

is equal to the limiting behavior of

$$\max_K \sup_{\theta \in \Theta_0(K,P)} \sum_{(j,\ell) \in K} \left(\frac{1}{\sqrt{p(x_j)}} \frac{1}{\sqrt{n}} \right.$$
$$\left. \times \sum_{1 \leq i \leq n} (-1)^{\ell-1}(Y_{\ell,i} - x_j'\theta)I\{X_i = x_j\} \right)_+^2$$
$$= \max_K \sup_{\theta \in \Theta_0(K,P)} \sum_{(j,\ell) \in K} \left(\frac{1}{\sqrt{p(x_j)}} \frac{1}{\sqrt{n}} \right.$$
$$\left. \times \sum_{1 \leq i \leq n} (-1)^{\ell-1}(Y_{\ell,i} - E_P[Y_{\ell,i}|X_i = x_j])I\{X_i = x_j\} \right)_+^2,$$

where $p(x_j) = P\{X_i = x_j\}$, the maximum over K is understood to be over all subsets of $\{1, \ldots, m\} \times \{1, 2\}$, and the supremum over the empty set is understood to be zero. The vector whose (j, ℓ) component is given by

$$\frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (-1)^{\ell-1}(Y_{\ell,i} - E_P[Y_{\ell,i}|X_i = x_j])I\{X_i = x_j\}$$

tends in distribution to a multivariate normal random variable. Let $Z_{j,\ell}(P)$ denote the (j, ℓ) component of this limiting multivariate normal random variable. It follows by the continuous mapping theorem that (16) tends in distribution to

$$(17) \qquad \max_K \sup_{\theta \in \Theta_0(K,P)} \sum_{(j,\ell) \in K} \left(\frac{1}{\sqrt{p(x_j)}} Z_{j,\ell}(P) \right)_+^2.$$

To determine for which α (17) is continuous at its $1 - \alpha$ quantile, first note that (17) is a convex function of $Z(P)$. By Theorem 11.1 of Davydov, Lifshits, and

Smorodina (1998), the distribution of (17) is continuous everywhere except possibly at zero, but

$$P\left\{\max_K \sup_{\theta \in \Theta_0(K,P)} \sum_{(j,\ell) \in K} \left(\frac{1}{\sqrt{p(x_j)}} Z_{j,\ell}(P)\right)_+^2 \leq 0\right\}$$

$$\leq P\left\{\left(\frac{1}{\sqrt{p(x_{j^*})}} Z_{j^*,\ell^*}(P)\right)_+^2 \leq 0\right\} \leq \frac{1}{2},$$

where j^*, ℓ^*, and θ^* are as in assumption (iii) above. Hence, (17) is continuous at its $1 - \alpha$ quantile for $\alpha < \frac{1}{2}$.

REMARK 2.3: A Tobit-like model is a special case of the above setup if we suppose further that $Y_{2,i} = Y_i^*$ and $Y_{1,i} = Y_i^*$ if $Y_i^* > 0$, and $Y_{2,i} = 0$ and $Y_{1,i} = -\infty$ (or some large negative number if there is a plausible lower bound on Y_i^*) if $Y_i^* \leq 0$.

REMARK 2.4: Our construction of critical values has used subsampling. Following Andrews (2000), it is possible to show that a naive bootstrap construction fails to approximate the distribution of (10) when $K = \Theta_0(P)$. It may still be the case that (11) is satisfied, but in simulations it seems to be too conservative in practice. Bugni (2007) showed that a suitably modified version of the bootstrap can be used to estimate the distribution of (10) when $K = \Theta_0(P)$. His approximation depends crucially on the structure of Example 2.1 and does not extend easily to more general models, but it is worthwhile to note that it can be used as an ingredient in Algorithm 2.1. Specifically, we may replace $\hat{c}_n(K, 1 - \alpha)$ with the $1 - \alpha$ quantile of his bootstrap approximation to the distribution of (10). It follows from the analysis of Bugni (2007) that these critical values will satisfy (11) under weak assumptions. Since they also satisfy (12), the conclusions of Theorem 2.1 follow.

REMARK 2.5: When the set of minimizers of $\hat{Q}_n(\theta)$ provides a consistent estimate of $\Theta_0(P)$, Chernozhukov, Hong, and Tamer (2007) proposed constructing confidence regions that satisfy (2) using a single-step method in which

$$(18) \qquad S_1 = \tilde{\Theta}_{0,n} = \arg\min_{\theta \in \Theta} \hat{Q}_n(\theta)$$

and critical values are given by (14); that is,

$$C_n = \{\theta \in \Theta : a_n \hat{Q}_n(\theta) \leq \hat{r}_n(\tilde{\Theta}_{0,n}, 1 - \alpha)\}.$$

Such an approach can be shown by example to fail to lead to confidence regions that satisfy (2) when the set of minimizers of $\hat{Q}_n(\theta)$ does not provide a consistent estimate of $\Theta_0(P)$. See Example 2.7 of Romano and Shaikh (2006) for details.

REMARK 2.6: When the set of minimizers of $\hat{Q}_n(\theta)$ does not provide a consistent estimate of $\Theta_0(P)$, Chernozhukov, Hong, and Tamer (2007) proposed constructing confidence regions that satisfy (2) using a single-step method in which

$$(19) \qquad S_1 = \hat{\Theta}_{0,n} = \{\theta \in \Theta : \hat{Q}_n(\theta) < \varepsilon_n\},$$

where ε_n is a positive sequence of constants tending to zero slowly. Because of this restriction on the rate at which ε_n tends to zero, they were able to show that

$$(20) \qquad P\{\Theta_0(P) \subseteq S_1\} \to 1.$$

The proof of Theorem 2.1 requires that the initial set S_1 be such that $\Theta_0(P) \subseteq S_1$, but one can allow for S_1 to be random provided that it satisfies (20) without affecting the argument in any way. Hence, using our results, it follows that this construction satisfies (2). Unfortunately, the specific choice of ε_n in finite samples is arbitrary and the confidence region resulting from application of their method may thus be very large or very small depending on the choice of ε_n. Our results provide a justification of iterating their procedure until a stopping criterion is met, thereby removing this arbitrariness, and produce typically smaller confidence regions while still maintaining the coverage requirement.

REMARK 2.7: It follows from the discussion in Remark 2.5 that there are no first-order differences between the confidence regions from our step-down procedure with S_1 given by (19) and those of Chernozhukov, Hong, and Tamer (2007). Even with such a delicate choice of S_1, we expect the iterative approach to perform better in finite samples. To this end, it is worthwhile to examine second-order differences. In Section A.3 of the Appendix, we show in the context of a simple example that our confidence region is smaller to second order than the one proposed by Chernozhukov, Hong, and Tamer (2007). In the example we consider, it is important to note that the set of minimizers of $\hat{Q}_n(\theta)$ provides a consistent estimate of $\Theta_0(P)$, so one could instead use a single-step procedure with S_1 given by (18). Compared with this procedure, our confidence region is not smaller to second order. We simply use the example to illustrate a phenomenon that we expect to persist even when the set of minimizers of $\hat{Q}_n(\theta)$ does not provide a consistent estimate of $\Theta_0(P)$.

2.2. *Uniform Consistency in Level*

We now provide conditions under which the set of θ values corresponding to the accepted hypotheses from Algorithm 2.1, $\mathcal{C}_n$, satisfies (3).

THEOREM 2.3: *Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P and let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$. Let $J_n(\cdot, P)$ denote the distribution of $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta)$ under P. Suppose $P \in \mathbf{P}$ and*

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

Then the set of θ values corresponding to the accepted hypotheses from Algorithm 2.1 using critical values given by (14), $\mathcal{C}_n$, satisfies (3).

PROOF: By Theorem 2.1, we have that

$$\text{FWER}_P \leq 1 - P\Big\{ \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \leq \hat{r}_n(\Theta_0(P), 1 - \alpha) \Big\}.$$

By Theorem 3.1(iv) of Romano and Shaikh (2008), it follows that

$$(22) \qquad \liminf_{n \to \infty} \inf_{P \in \mathbf{P}} P\Big\{ \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \leq \hat{r}_n(\Theta_0(P), 1 - \alpha) \Big\} \geq 1 - \alpha.$$

Thus,

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}} \text{FWER}_P \leq \alpha.$$

The asserted claim now follows immediately from Lemma 2.1.   Q.E.D.

The intuition behind condition (21) is as follows. Lemma A.1 shows under weak conditions that the subsampling estimator of the distribution of $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta)$ approximates $J_b(x, P)$ well uniformly over both $x \in \mathbf{R}$ and $P \in \mathbf{P}$. Condition (21) implies that critical values from $J_b(x, P)$ are no smaller than those from $J_n(x, P)$. Hence, under this assumption, Theorem 3.1 of Romano and Shaikh (2008) implies that subsampling behaves well over $P \in \mathbf{P}$ in the sense that (22) holds.

We now apply Theorem 2.3 to construct confidence regions that satisfy the coverage requirement (3) for the two examples considered in Section 2.

EXAMPLE 2.3—Moment Inequalities: Recall the setup of Example 2.1. We will now use Theorem 2.3 to show that for this example the set of θ values corresponding to the accepted hypotheses from Algorithm 2.1, $\mathcal{C}_n$, satisfies (3) for a large class of distributions $\mathbf{P}$. To this end, let $a_n = n$, let $\mathbf{P}$ be such that (i)

$$\{g_j(\cdot, \theta) : 1 \leq j \leq m, \theta \in \Theta\} \quad \text{is } P\text{-Donsker and pre-Gaussian}$$

uniformly in $P \in \mathbf{P}$ and (ii) Θ is compact with respect to the metric

$$\bar{\rho}(\theta, \theta') = \sup_{P \in \mathbf{P}} \max_{1 \le j \le m} \rho_{P,j}(\theta, \theta'),$$

where $\rho_{P,j}^2(\theta, \theta') = E_P[(g_j(X, \theta) - g_j(X, \theta'))^2]$. To rule out degenerate situations, assume further that (iii) there exists $\varepsilon > 0$ such that for each $P \in \mathbf{P}$ there exist $1 \le j^* \le m$ and $\theta^* \in \Theta_0(P)$ such that $E_P[g_{j^*}(X_i, \theta^*)] = 0$ and $\mathrm{Var}_P[g_{j^*}(X_i, \theta^*)] \ge \varepsilon$. Assumption (i) is again known to hold provided that the class of functions is not too large; for general results to this end and numerous applications, see van der Vaart and Wellner (1996). In the Appendix, we verify that the required condition (21) holds under these assumptions.

EXAMPLE 2.4—Regression With Interval Outcomes: Recall the setup of Example 2.2. As argued there, the structure of this example is similar to that of Example 2.3. Since we provide details in the case of Example 2.3 above, we do not do so here.

2.3. *Confidence Regions for Functions of the Identified Set*

In this section, we consider the problem of constructing sets that satisfy (5) and (6). Let $f : \Theta \to \Lambda$ be given. Our construction again relies on equivalence with an appropriate multiple testing problem, but in this case the family of null hypotheses is given by

(23) $H_\lambda : \lambda \in \Lambda_0(P) \quad \text{for} \quad \lambda \in \Lambda,$

where $\Lambda_0(P)$ is defined by (4). The alternative hypotheses are understood to be

$$K_\lambda : \lambda \notin \Lambda_0(P) \quad \text{for} \quad \lambda \in \Lambda.$$

As before, it suffices to consider the problem of testing this family of null hypotheses in a way that controls the FWER_P at level α.

For $\lambda \in \Lambda$, let $f^{-1}(\lambda) = \{\theta \in \Theta : f(\theta) = \lambda\}$. Note that

$$\lambda \in \Lambda_0(P) \iff \exists \theta \in f^{-1}(\lambda) \quad \text{s.t. } Q(\theta, P) = 0$$
$$\implies \inf_{\theta \in f^{-1}(\lambda)} Q(\theta, P) = 0.$$

This suggests a natural test statistic for each of these null hypotheses H_λ:

(24) $\inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta),$

where $a_n \hat{Q}_n(\theta)$ is the test statistic used earlier to test the null hypothesis that $Q(\theta, P) = 0$.

We may now proceed as before, but with this test statistic in place of our earlier test statistic $a_n \hat{Q}_n(\theta)$. For $K \subseteq \Lambda$, let $\hat{c}_n^f(K, 1 - \alpha)$ be an estimator of the $1 - \alpha$ quantile of distribution of

$$\sup_{\lambda \in K} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta)$$

and consider the following modification of Algorithm 2.1.

ALGORITHM 2.2:
1. Let $S_1 = \Lambda$. If $\sup_{\lambda \in S_1} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}_n^f(S_1, 1 - \alpha)$, then accept all H_λ and stop; otherwise, set $S_2 = \{\lambda \in \Lambda : \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}_n^f(S_1, 1 - \alpha)\}$ and continue.
    ⋮
j. If $\sup_{\lambda \in S_j} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}_n^f(S_j, 1 - \alpha)$, then accept all H_λ with $\lambda \in S_j$ and stop; otherwise, set $S_{j+1} = \{\lambda \in \Lambda : \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}_n^f(S_j, 1 - \alpha)\}$ and continue.
    ⋮

We now provide conditions under which the set of θ values corresponding to accepted hypotheses from Algorithm 2.2 leads to confidence regions that satisfy (5) and (6). For $K \subseteq \Lambda$ and $\alpha \in (0, 1)$, let

$$(25) \quad \hat{r}_n^f(K, 1 - \alpha)$$

$$= \inf \left\{ x : \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I \left\{ \sup_{\lambda \in K} \inf_{\theta \in f^{-1}(\lambda)} a_b \hat{Q}_{n,b,i}(\theta) \leq x \right\} \geq 1 - \alpha \right\}.$$

THEOREM 2.4: *Let* $X_i, i = 1, \ldots, n$, *be an i.i.d. sequence of random variables with distribution* P *and let* $b = b_n < n$ *be a sequence of positive integers tending to infinity, but satisfying* $b/n \to 0$. *Let* $J_n(\cdot, P)$ *denote the distribution of* $\sup_{\lambda \in \Lambda_0(P)} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta)$ *under* P. *Let* $\mathcal{C}_n^f$ *denote the set of* θ *values corresponding to accepted hypotheses from Algorithm 2.2 when* $\hat{c}_n^f(K, 1 - \alpha)$ *is given by* (25).

  (i) *Suppose* $J_n(\cdot, P)$ *converges in distribution to* $J(\cdot, P)$ *and that* $J(\cdot, P)$ *is continuous at its smallest* $1 - \alpha$ *quantile. Then* $\mathcal{C}_n^f$ *satisfies* (5).

  (ii) *Suppose* $P \in \mathbf{P}$ *and*

$$(26) \quad \limsup_{n \to \infty} \sup_{P \in \mathbf{P}} \sup_{x \in \mathbf{R}} \{J_b(x, P) - J_n(x, P)\} \leq 0.$$

Then $\mathcal{C}_n^f$ *satisfies* (6).

The proof follows immediately from the arguments given in Sections 2.1 and 2.2.

We now provide a simple illustration of the use of Theorem 2.4.

EXAMPLE 2.5: Let (X_i, Y_i), $i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P on $\mathbf{R}^2$. The parameter of interest, θ_0, is known to satisfy $\theta_{0,1} \geq \mu_X(P)$ and $\theta_{0,2} \geq \mu_Y(P)$. The identified set is therefore given by $\Theta_0(P) = \{\theta \in \mathbf{R}^2 : \theta_1 \geq \mu_X(P) \text{ and } \theta_2 \geq \mu_Y(P)\}$. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = (\mu_X(P) - \theta_1)_+^2 + (\mu_Y(P) - \theta_2)_+^2.$$

The sample analog of $Q(\theta, P)$ is given by $\hat{Q}_n(\theta) = (\bar{X}_n - \theta_1)_+^2 + (\bar{Y}_n - \theta_2)_+^2$. Suppose the object of interest is the projection of $\Theta_0(P)$ onto its first component rather than the entire set $\Theta_0(P)$; that is, the object of interest is $\Lambda_0(P) = f(\Theta_0(P))$, where $f : \mathbf{R}^2 \to \mathbf{R}$ is defined by $f(\theta) = \theta_1$ instead of $\Theta_0(P)$. Note that $\Lambda_0(P)$ is simply $\{\theta_1 \in \mathbf{R} : \theta_1 \geq \mu_X(P)\}$.

First consider the problem of constructing sets that satisfy (5). Let $a_n = n$ and suppose P is such that $\sigma_X^2(P)$ exists. Assume without loss of generality that $\mu_X(P) = 0$. Then

$$\sup_{\theta_1 \in \Lambda_0(P)} \inf_{f^{-1}(\theta_1)} a_n \hat{Q}_n(\theta) = \sup_{\theta_1 \geq 0} \inf_{\theta_2 \in \mathbf{R}} n(\bar{X}_n - \theta_1)_+^2 + n(\bar{Y}_n - \theta_2)_+^2$$

$$= n(\bar{X}_n)_+^2 \xrightarrow{\mathcal{L}} (\sigma_X(P)Z)_+^2,$$

where Z is a standard normal random variable. It now follows from Theorem 2.4(i) that the set of θ values corresponding to accepted hypotheses from Algorithm 2.2 when $\hat{c}_n^f(K, 1 - \alpha)$ is given by (25), $\mathcal{C}_n^f$, satisfies (5).

Now consider the problem of constructing sets that satisfy (6). As before, let $a_n = n$ and let $\mathbf{P}$ be a set of distributions for which the marginal distribution of X satisfies

$$(27) \qquad \lim_{\lambda \to \infty} \sup_{P \in \mathbf{P}} E_P\left[\frac{|X - \mu(P)|^2}{\sigma^2(P)} I\left\{\frac{|X - \mu(P)|}{\sigma(P)} > \lambda\right\}\right] = 0.$$

After noting that $\sup_{\theta_1 \in \Lambda_0(P)} \inf_{f^{-1}(\theta_1)} a_n \hat{Q}_n(\theta)$ is simply $n(\bar{X}_n)_+^2$, it is straightforward to apply Lemma 11.4.1 of Lehmann and Romano (2005) to show that (26) holds. Therefore, it follows from Theorem 2.4(ii) that the set of θ values corresponding to accepted hypotheses from Algorithm 2.2 when $\hat{c}_n^f(K, 1 - \alpha)$ is given by (25), $\mathcal{C}_n^f$, satisfies (6).

REMARK 2.8: Given a confidence region for the identified set $\mathcal{C}_n$, one construction of a confidence region for a function of the identified set is the image

of $\mathcal{C}_n$ under the function of interest. Such a construction will typically be conservative in the sense that the coverage probability will exceed the nominal level.

3. EMPIRICAL ILLUSTRATION

In this section, we use the techniques developed above to examine the impact of top-coding outcomes on the inferences that can be made about the parameters of a linear regression. By top-coding a random variable, we mean the practice of recording the realization of the random variable if and only if it is below a certain threshhold. This model is a special case of our Example 2.2, and so the theory developed above applies here under the appropriate assumptions. A similar empirical example can be found in Chernozhukov, Hong, and Tamer (2004).

The motivation for our exercise stems from the following observation. To study changes in the wage structure and earnings inequality, researchers often regress the logarithm of hourly wages on various demographic characteristics. Data sets used for this purpose invariably top-code wages for reasons of confidentiality. One approach to deal with the top-coding of wages is to replace all of the top-coded outcomes with a common value. In practice, this common value is often taken to be a scalar multiple of the threshhold. This approach is justified theoretically under the assumption that the distribution of wages conditional on top-coding is distributed as a Pareto random variable. See, for example, Katz and Autor (1999), wherein the scalar used for this purpose is taken to be 1.5. Of course, we do not wish to impose any parametric assumptions.

To examine this issue, we begin with a sample of observations from the Annual Demographic Supplement of the Current Population Survey for the year 2000. For each individual in the survey, the survey records a variety of demographic variables as well as information on wages and salaries. We select observations with the following demographic characteristics: (1) race is white; (2) age is between 20 and 24 years; (3) at least college graduates; (4) primary source of income is wages and salaries; (5) worked at least 2 hours per week on average. There are 305 such observations, none of which suffers from top-coding of wages and salaries. We treat this sample of individuals as the distribution of the observed data P and draw an i.i.d. sample of $n = 1000$ observations from this P. We will analyze these data both for the benchmark case of no top-coding and for cases in which some amount of top-coding has been artificially imposed on the data.

Recall the setup of Example 2.2. To allow for graphical illustration of the confidence regions, we consider only a model in which $k = 2$; specifically, we take $X_i = (1, D_i)$, where D_i is 1 if the sex is female and 0 otherwise. The latent outcome variable $Y_i^* = \log(\text{wage}_i^*/H_i)$, where wage_i^* is total wages and salaries, which is possibly unobserved in the presence of top-coding, and H_i is total

hours worked. We assume that wage$_i^*$ is bounded above by $\overline{\text{wage}} = \$10^8$. In the benchmark case in which there is no top-coding, we will let $Y_{1,i} = Y_{2,i} = Y_i^*$. In the cases in which there is some top-coding, let wage be the threshhold above which wages are not observed. Define $Y_{1,i} = Y_{2,i} = Y_i^*$ if wage$_i^* \leq$ wage; otherwise, let $Y_{1,i} = \underline{Y}_i = \log(\text{wage}/H_i)$ and $Y_{2,i} = \overline{Y}_i = \log(\overline{\text{wage}}/H_i)$.

Below we will construct confidence regions of level $1 - \alpha = 0.95$ for the identified set for each of three different scenarios. For the sake of completeness, we will also construct confidence regions for identifiable parameters, as described in Romano and Shaikh (2008). More specifically, following Romano and Shaikh (2008), we consider

$$\left\{ \theta \in \mathbf{R}^2 : a_n \hat{Q}_n(\theta) \leq \hat{r}_n(\{\theta\}, 1 - \alpha) \right\},$$

where $\hat{r}_n(\{\theta\}, 1 - \alpha)$ is given by (14). We will compare the inferences that can be drawn from these confidence regions with those that can be drawn from regressing Y_i^a on X_i, where, in the benchmark case of no top-coding, $Y_i^a = Y_i^*$, and, in cases with top-coding, $Y_i^a = Y_i^*$ if wage$_i^* \leq$ wage and $Y_i^a = 1.5 \times \underline{Y}_i$ otherwise.

Before proceeding, we discuss some computational details. First, consider the choice of b. In practice, one would like to use a data-dependent subsample size; see Politis, Romano, and Wolf (1999) for a review of several algorithms for choosing the subsample size in this way. For the purposes of this exercise, however, we use the same subsample size, $b = 30$, in each of the constructions. As a result, differences among the confidence regions below are not driven by variation in the choice of subsample size. Note that the results below remain similar for subsample sizes between 20 and 40, the range of subsample sizes for which the simulation results in the following section suggest that the procedure behaves well in finite samples. Second, when computing critical values, we also used an approximation as described in Remark 2.1 with $B = 200$ because $\binom{n}{b}$ is too large to compute critical values exactly. Finally, following the discussion in Remark 2.5, in the first step of Algorithm 2.1, we let

$$S_1 = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 1000\}.$$

The results below remain similar for much larger choices of S_1.

We first consider the case in which there is no top-coding. Algorithm 2.1 converged after 11 steps and the confidence region for the identified set is given by

$$\mathcal{C}_n = S_{11} = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 0.0055\}.$$

We also regress Y_i^a on X_i and obtain a Wald-style confidence region of the form

$$\{\theta \in \mathbf{R}^2 : (\hat{\theta}_n - \theta)' \hat{\Sigma}_n^{-1} (\hat{\theta}_n - \theta) \leq 5.99\},$$
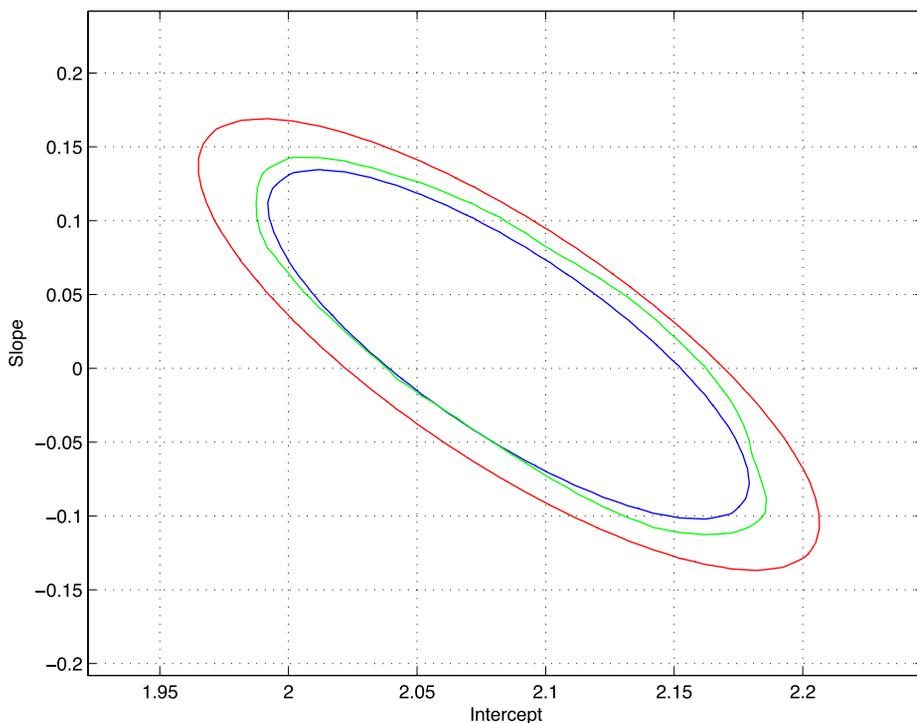
FIGURE 1.—Confidence regions with no top-coding and $\overline{wage} = \$10^8$: green = confidence region for identifiable parameters, red = confidence region for identified set, blue = Wald-style confidence region.

where $\hat{\Sigma}_n$ is the usual heteroskedasticity-robust estimator of the variance of $\hat{\theta}_n$. These two confidence regions together with the confidence region for identifiable parameters are displayed in Figure 1. Since the true P that generates the data is known, it is also possible to calculate the identified set, which in this case is a singleton. It is given by $\Theta_0(P) = \{(2.047, 0.042)\}$. As one would expect, in this instance all three confidence regions are of similar shape and size. The largest is the confidence region for the identified set and the smallest is the Wald-style confidence region. The confidence region for identifiable parameters is contained strictly within the confidence region for the identified set.

Next, we consider a case in which there is some amount of top-coding and repeat the exercise above. For concreteness, we choose $\underline{wage} = \$41,000$, which corresponds to 5% of the population being subject to top-coding. Algorithm 2.1 converged after 12 steps and the confidence region for the identified set is given by

$$\mathcal{C}_n = S_{12} = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 0.0182\}.$$
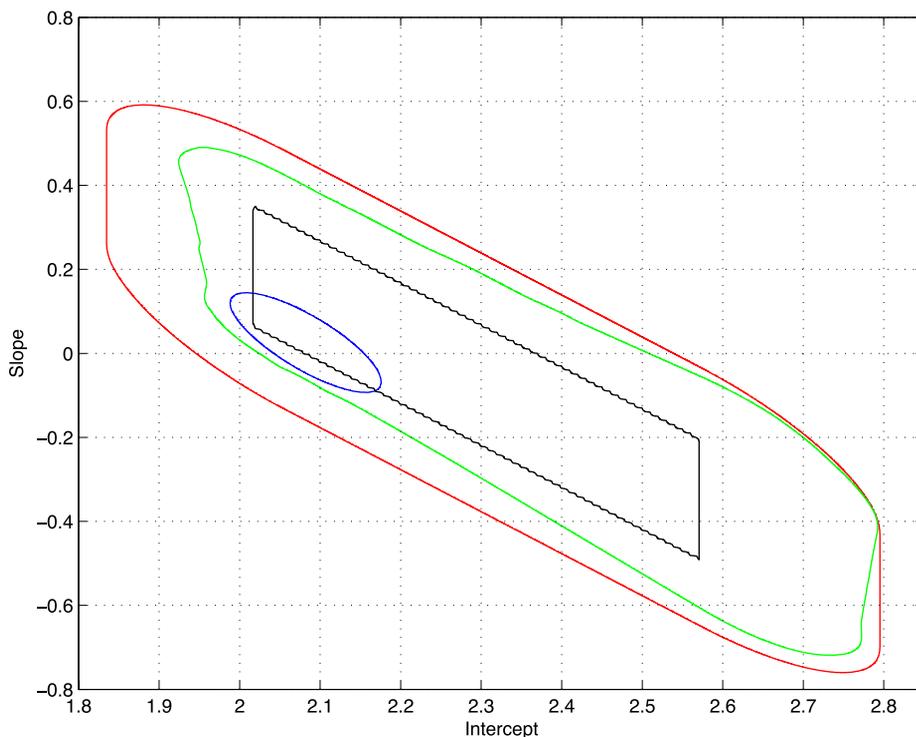
FIGURE 2.—Confidence regions with 5% top-coding and $\overline{\text{wage}} = \$10^8$: green = confidence region for identifiable parameters, red = confidence region for identified set, blue = Wald-style confidence region, black = identified set.

The resulting confidence regions are displayed in Figure 2. Again, we may also calculate the identified set, which is no longer a singleton due to top-coding. It is given by

$$\left\{ \theta \in \mathbf{R}^2 : E\{Y_{1,i}|D_i = 0\} \le \theta_1 \le E\{Y_{2,i}|D_i = 0\}, \right.$$
$$\left. E\{Y_{1,i}|D_i = 1\} \le \theta_1 + \theta_2 \le E\{Y_{2,i}|D_i = 1\} \right\}$$

and is therefore a parallelogram. This set is also displayed in Figure 2. Both the confidence region for the identified set and the confidence region for identifiable parameters contain the identified set, but, as before, the confidence region for identifiable parameters is contained strictly within the confidence region for the identified set. The Wald-style confidence region, though still the smallest, covers only a small portion of the identified set. As a result, inferences based on the Wald-style confidence region might be very misleading if the assumptions used to achieve identification are not correct.

To make this point more forcefully, we carry out the same exercise for the case in which there is even more top-coding. Specifically, we reduce $\underline{wage}$ to $35,000, which corresponds to 10% of the population being subject to top-coding. Algorithm 2.1 converged after 9 steps and the confidence region for the identified set is given by

$$\mathcal{C}_n = S_9 = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 0.0361\}.$$

The resulting confidence regions along with the identified set are displayed in Figure 3. The qualitative features of this figure are the same as before, except now the Wald-style confidence region covers an even smaller portion of the identified set, and so inferences based on it may be even more misleading.

Of course, so far we have assumed a very generous upper bound on annual wages and salaries of $\overline{wage} = \$10^8$. To assess how sensitive the qualitative results described above are to the value of $\overline{wage}$, we reexamine the previous case
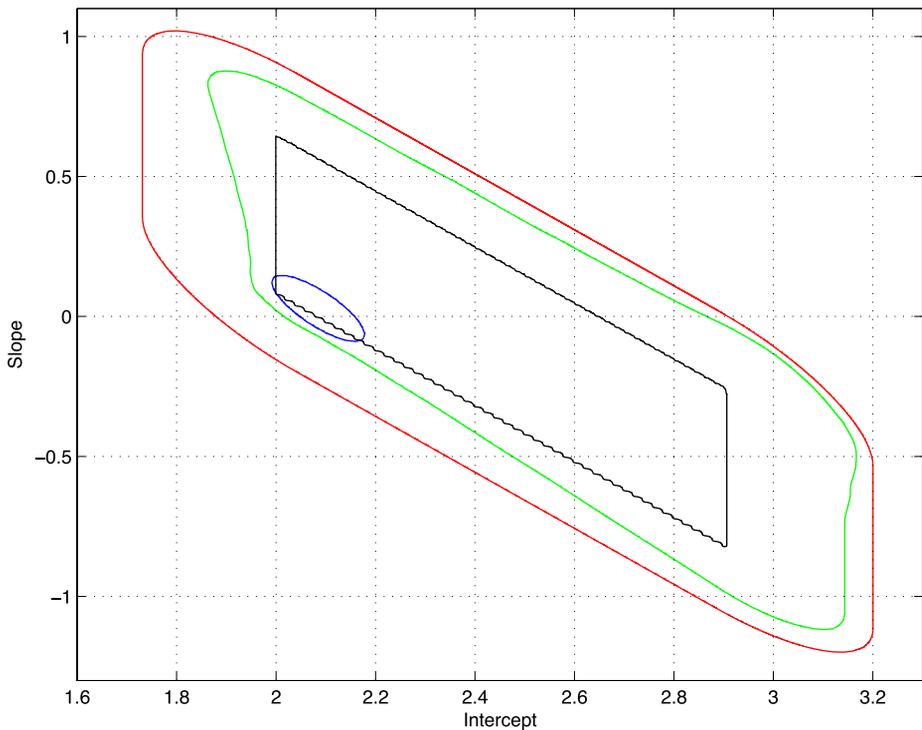
FIGURE 3.—Confidence regions with 10% top-coding and $\overline{wage} = \$10^8$: green = confidence region for identifiable parameters, red = confidence region for identified set, blue = Wald-style confidence region, black = identified set.
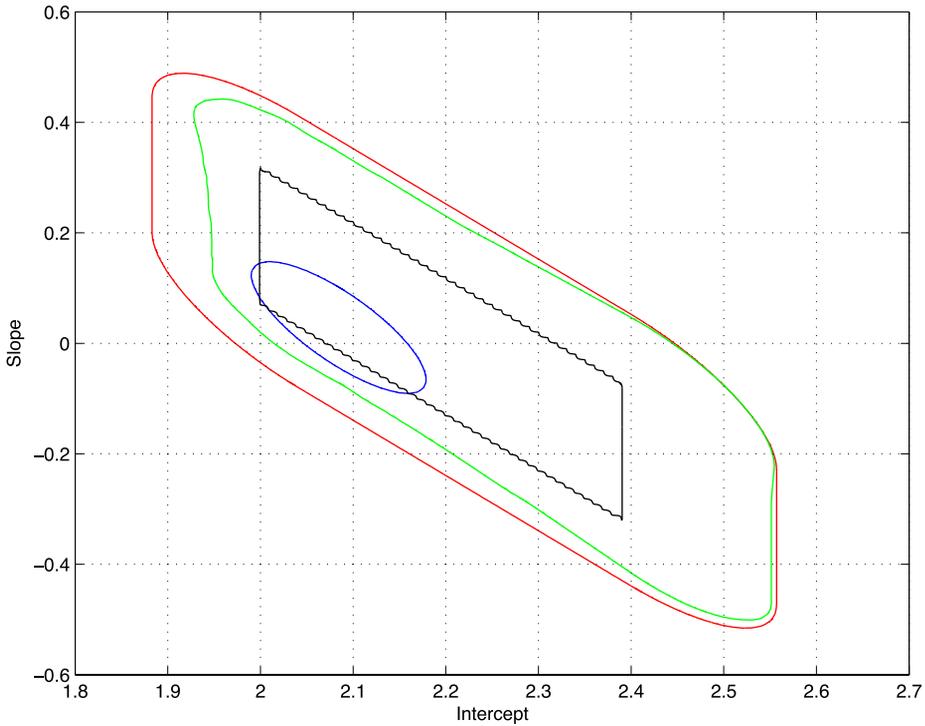
FIGURE 4.—Confidence regions with 5% top-coding and $\overline{\text{wage}} = \$10^6$: green = confidence region for identifiable parameters, red = confidence region for identified set, blue = Wald-style confidence region, black = identified set.

in which 10% of the population is subject to top-coding with the much lower value of $\overline{\text{wage}} = \$10^6$. Algorithm 2.1 converged after 12 steps and the confidence region for the identified set is then given by

$$\mathcal{C}_n = S_{12} = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 0.0094\}.$$

The confidence regions from this exercise along with the identified set are displayed in Figure 4. Again, the qualitative features of this figure are the same as before, but, as one would expect, the identified set is smaller than before. This suggests that in applications, the choice of $\overline{\text{wage}}$ is important, as it will noticeably impact the sharpness of inferences in such a setting.

4. SIMULATION RESULTS

In this section, we shed some light on the finite-sample behavior of our step-down procedure via a small simulation study. For the simulation study, we

use the final specification of the empirical illustration in which $\overline{wage} = \$10^6$. As in the empirical illustration, the sample size n is 1000 and $\alpha = 0.05$. Following the discussion in Remark 2.5, in the first step of Algorithm 2.1, we set $S_1 = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \le 1000\}$. To assess the sensitivity of our procedure to the choice of subsample size, we consider values of b in $\{15, 20, \ldots, 85\}$. Finally, as described in Remark 2.1, we approximate the critical values with $B = 200$.

For each of 100 simulations, we compute the variables (i) j^*, the iteration at which Algorithm 2.1 converged, (ii) $\hat{c}_n(S_{j^*}, 1 - \alpha)$, the critical value that defines $\mathcal{C}_n$, (iii) $\hat{c}_n(S_{j^*-1}, 1 - \alpha)$, (iv) $\sup_{\Theta_0(P)} \hat{Q}_n(\theta)$, and (v) $I\{\Theta_0(P) \subseteq \mathcal{C}_n\}$.

In Table I, we present, for each value of b, (i) the average number of iterations needed for Algorithm 2.1 to converge, that is, the average value of $\hat{j}^*$, and (ii) the simulated probability that the identified set is covered by $\mathcal{C}_n$. The simulation results show that the average number of iterations increases with the subsample size, but it is typically between 7 and 10. The simulation results also show that the coverage probabilities are close to the nominal level, $1 - \alpha$, for values of b ranging from 20 to 40.

In Table II, we present, for each value of b, (i) the mean of $\hat{c}_n(S_{j^*}, 1 - \alpha)$, (ii) the mean of $\hat{c}_n(S_{j^*-1}, 1 - \alpha)$, and (iii) the simulated $1 - \alpha$ quantile of $\sup_{\Theta_0(P)} \hat{Q}_n(\theta)$. We label the third column "ideal" because it represents the best possible critical value. Of course, it is infeasible, since it depends on P, which is typically unknown. Fortunately, the simulation results show that for values of b between 20 and 40, $\hat{c}_n(S_{j^*}, 1 - \alpha)$ is close to this ideal value. The simulation results also allow for a comparison with single-step procedures. To see this, recall

TABLE I

SIMULATION RESULTS

b	Coverage Probability	Average Number of Iterations
15	0.98	6.99
20	0.94	7.02
25	0.97	7.64
30	0.91	7.98
35	0.98	8.04
40	0.96	8.14
45	0.97	8.87
50	1.00	8.98
55	0.98	9.08
60	0.99	9.20
65	0.99	9.77
70	0.98	9.95
75	0.99	10.02
80	1.00	10.08
85	1.00	10.24

TABLE II

SIMULATION RESULTS—CRITICAL VALUES

b	Avg. $\hat{c}_n(S_{j*}, 1-\alpha)$	Avg. $\hat{c}_n(S_{j*-1}, 1-\alpha)$	Ideal
15	0.0063	0.0066	0.0057
20	0.0069	0.0073	0.0066
25	0.0074	0.0077	0.0057
30	0.0076	0.0078	0.0077
35	0.0079	0.0082	0.0052
40	0.0084	0.0088	0.0064
45	0.0084	0.0086	0.0062
50	0.0089	0.0092	0.0047
55	0.0090	0.0093	0.0056
60	0.0093	0.0096	0.0060
65	0.0097	0.0099	0.0052
70	0.0098	0.0100	0.0071
75	0.0102	0.0105	0.0054
80	0.0103	0.0105	0.0058
85	0.0107	0.0109	0.0052

that for a single-step procedure to lead to a smaller confidence region than our step-down procedure, we would have to choose S_1 to be smaller than

$$\{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq \hat{c}_n(S_{j*-1}, 1-\alpha)\}.$$

Since $\hat{c}_n(S_{j*-1}, 1-\alpha)$ is typically very small in the simulation results, we would have to choose S_1 to be very small as well for the single-step procedure to lead to a smaller confidence region than our step-down procedure. Moreover, such a confidence region, though smaller than ours, may have poor coverage probabilities in finite samples.

APPENDIX

A.1. *Auxiliary Results*

LEMMA A.1: *Let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables with distribution P. Denote by $J_n(\cdot, P)$ the distribution of the statistic $\tau_n(\hat{\theta}_n - \theta(P))$. Suppose $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ is a symmetric function of its arguments and that $\hat{\theta}_n \in S$. Let $\mathcal{V}$ be a Vapnik–Chervonenkis (VC) class of subsets of S with VC index v and assume that $\mathcal{V}$ is permissible. For $0 < b < n$, let $N_n = \binom{n}{b}$ and let $k_n = \lfloor \frac{n}{b} \rfloor$. Then, for any $\varepsilon > 0$, we have that*

$$(28) \quad P\left\{ \sup_{V \in \mathcal{V}} \left| \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\{\tau_b(\hat{\theta}_{n,b,i} - \theta(P)) \in V\} - J_b(V, P) \right| > \varepsilon \right\}$$

is bounded above by

$$(29) \qquad \frac{1}{\varepsilon}\left(\left(\sqrt{\frac{2}{k_n}} \vee 4\sqrt{\frac{2\log(8k_n^v)}{k_n}}\right) \wedge 1\right)$$

$$+ \frac{32k_n^v}{\varepsilon}\sqrt{\frac{2\pi}{k_n}}\left[\Phi\left(\frac{4}{\sqrt{k_n}}\right) - \Phi\left(\left(\frac{1}{\sqrt{k_n}} \vee \sqrt{2\log(8k_n^v)}\right) \wedge \frac{4}{\sqrt{k_n}}\right)\right],$$

where $\Phi(\cdot)$ *is the standard normal distribution. We also have that for any* $0 < \delta < 1$, (28) *is bounded above by*

$$(30) \qquad \frac{\delta}{\varepsilon} + \frac{1}{\varepsilon}8k_n^v\exp\left\{-\frac{k_n\delta^2}{32}\right\}$$

whenever $k_n\delta^2 \geq 2$.

PROOF: For $V \in \mathcal{V}$ define

$$S_n(V, P; X_1, \ldots, X_n)$$
$$= \frac{1}{k_n}\sum_{1 \leq i \leq k_n} I\left\{\tau_b\big(\hat{\theta}_b\big(X_{b(i-1)+1}, \ldots, X_{bi}\big) - \theta(P)\big) \in V\right\} - J_b(V, P).$$

Denote by $\mathcal{S}_n$ the symmetric group with n elements. Note that using this notation, we may rewrite

$$\frac{1}{N_n}\sum_{1 \leq i \leq N_n} I\left\{\tau_b(\hat{\theta}_{n,b,i} - \theta(P)) \in V\right\} - J_b(V, P)$$

as

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

Note further that

$$\sup_{V \in \mathcal{V}}|Z_n(V, P; X_1, \ldots, X_n)| \leq \frac{1}{n!}\sum_{\pi \in \mathcal{S}_n}\sup_{V \in \mathcal{V}}\big|S_n\big(V, P; X_{\pi(1)}, \ldots, X_{\pi(n)}\big)\big|,$$

which is a sum of $n!$ identically distributed random variables. Let $\varepsilon > 0$ be given. It follows that

$$(31) \qquad P\left\{\sup_{V \in \mathcal{V}}|Z_n(V, P; X_1, \ldots, X_n)| > \varepsilon\right\}$$

$$\leq P\left\{\frac{1}{n!}\sum_{\pi \in \mathcal{S}_n}\sup_{V \in \mathcal{V}}\big|S_n\big(V, P; X_{\pi(1)}, \ldots, X_{\pi(n)}\big)\big| > \varepsilon\right\}.$$

Using Markov's inequality, the right-hand side of (31) can be bounded by

$$(32) \qquad \frac{1}{\varepsilon} E\left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| \right\}$$

$$(33) \qquad = \frac{1}{\varepsilon} \int_0^1 P\left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| > u \right\} du.$$

Recall that the generalized Glivenko–Cantelli theorem asserts that

$$P\left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| > u \right\}$$

is bounded above by $8k_n^v \exp\{-k_n u^2/32\}$ whenever $k_n u^2 \geq 2$ and by 1 otherwise. It follows that (33) is bounded above by

$$\frac{1}{\varepsilon} \int_0^{(\sqrt{2/k_n} \vee 4\sqrt{(2\log(8k_n^v))/k_n}) \wedge 1} 1 \, du$$

$$+ \frac{1}{\varepsilon} \int_{(\sqrt{2/k_n} \vee 4\sqrt{(2\log(8k_n^v))/k_n}) \wedge 1}^1 8k_n^v \exp\left\{ -\frac{k_n u^2}{32} \right\} du.$$

Evaluating this last expression yields the bound (29). To establish (30), note that for any $0 < \delta < 1$, we have that

$$(34) \qquad E\left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| \right\}$$

$$\leq \delta + P\left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| > \delta \right\}.$$

The result (30) now follows immediately by using an exponential inequality as found in the proof of the Glivenko–Cantelli theorem for VC classes (see Section 2 of Pollard (1984)) to bound the second term on the right-hand side in (34). *Q.E.D.*

LEMMA A.2: *Let F and F_n, $n \geq 1$, be distribution functions on* **R**. *Suppose $F_n(x) \to F(x)$ for all $x \geq 0$ and that F is continuous on $(0, \infty)$. Then*

$$\sup_{x \geq 0} |F_n(x) - F(x)| \to 0.$$

For brevity, we omit the proof of this generalization of Polya's theorem.

A.2. *Technical Details for Example 2.1*

Note that we may write

$$(35) \qquad \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) = \sup_{\theta \in \Theta_0(P)} \sum_{1 \le j \le m} \left(Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)] \right)_+^2,$$

where

$$(36) \qquad Z_{n,j}(\theta, P) = \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \left(g_j(X_i, \theta) - E_P[g_j(X_i, \theta)] \right).$$

Let $0 > \lambda_n \to 0$, but so slowly that $\sqrt{n}\lambda_n \to -\infty$ and, for $K \subseteq \{1, \ldots, m\}$, define

$$\Theta_n'(K, P) = \left\{ \theta \in \Theta_0(P) : \lambda_n < E_P[g_j(X_i, \theta)] \iff j \in K \right\},$$

$$\Theta_n(K, P) = \left\{ \theta \in \Theta_0(P) : \lambda_n < E_P[g_j(X_i, \theta)] \text{ for all } j \in K \right\},$$

$$\Theta_0(K, P) = \left\{ \theta \in \Theta_0(P) : E_P[g_j(X_i, \theta)] = 0 \text{ for all } j \in K \right\}.$$

Note that

$$\Theta_0(P) = \bigcup \{\Theta_0(K, P) : K \subseteq \{1, \ldots, m\}\}$$

$$= \bigcup \{\Theta_n'(K, P) : K \subseteq \{1, \ldots, m\}\}$$

and adopt the convention that the sum over the empty set and the supremum over the empty set are zero. Hence, (35) can be bounded from below as

$$(37) \qquad \max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_0(K, P)} \sum_{1 \le j \le m} \left(Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)] \right)_+^2$$

$$\ge \max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_0(K, P)} \sum_{j \in K} \left(Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)] \right)_+^2$$

$$= \max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_0(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))_+^2.$$

On the other hand, (35) can be bounded from above as

$$\max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_n'(K, P)} \sum_{1 \le j \le m} \left(Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)] \right)_+^2$$

$$\le \max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_n'(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))_+^2$$

$$+ \max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_n'(K, P)} \sum_{j \notin K} \left(Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)] \right)_+^2$$

$$\leq \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_n(K,P)} \sum_{j \in K} (Z_{n,j}(\theta, P))_+^2$$

$$(38) \qquad + \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_n'(K,P)} \sum_{j \notin K} \left(Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)]\right)_+^2$$

$$(39) \qquad = \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_n(K,P)} \sum_{j \in K} (Z_{n,j}(\theta, P))_+^2 + o_P(1).$$

To see the equality (39), note that

$$\sup_{\theta \in \Theta_n'(K,P)} \sum_{j \notin K} \left(Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)]\right)_+^2$$

$$\leq m \left(\sup_{\theta \in \Theta_0(P)} \max_{1 \leq j \leq m} |Z_{n,j}(\theta, P)| + \sqrt{n} \lambda_n \right)_+^2.$$

By assumption (i),

$$(40) \qquad \sup_{\theta \in \Theta_0(P)} \max_{1 \leq j \leq m} |Z_{n,j}(\theta, P)| = O_P(1).$$

It thus follows from the assumption that $\sqrt{n} \lambda_n \to -\infty$ that (38) tends in probability to zero, which in turn implies the equality (39).

Next, we argue that

$$(41) \qquad \sup_{\theta \in \Theta_n(K,P)} \sum_{j \in K} (Z_{n,j}(\theta, P))_+^2 - \sup_{\theta \in \Theta_0(K,P)} \sum_{j \in K} (Z_{n,j}(\theta, P))_+^2 \xrightarrow{P} 0.$$

Since $\Theta_0(K, P) \subseteq \Theta_n(K, P)$, the left-hand side of (41) is bounded from below by zero. It therefore suffices to show that (41) is bounded from above by zero in probability. To this end, note that the left-hand side of (41) is bounded from above by

$$(42) \qquad \sum_{j \in K} (Z_{n,j}(\hat{\theta}_n, P))_+^2 + \varepsilon_n - \sup_{\theta \in \Theta_0(K,P)} \sum_{j \in K} (Z_{n,j}(\theta, P))_+^2,$$

where $0 < \varepsilon_n \to 0$ and $\hat{\theta}_n \in \Theta_n(K, P)$ are such that

$$\sup_{\theta \in \Theta_n(K,P)} \sum_{j \in K} (Z_{n,j}(\theta, P))_+^2 < \sum_{j \in K} (Z_{n,j}(\hat{\theta}_n, P))_+^2 + \varepsilon_n.$$

It is therefore enough to show that (42) is bounded from above in probability by zero. To see this, suppose by way of contradiction that there exists $\delta > 0$ such that

$$P\left\{\sum_{j\in K}(Z_{n,j}(\hat{\theta}_n, P))^2_+ - \sup_{\theta\in\Theta_0(K,P)}\sum_{j\in K}(Z_{n,j}(\theta, P))^2_+ > \delta\right\} \not\to 0.$$

By assumption (ii), it follows that there exists a subsequence n_k and a $\hat{\theta} \in \Theta_0(K, P)$ such that $\rho_{P,j}(\hat{\theta}_{n_k}, \hat{\theta}) \to 0$ for all $j \in K$ and

$$(43) \qquad P\left\{\sum_{j\in K}\left(Z_{n_k,j}(\hat{\theta}_{n_k}, P)\right)^2_+ - \sum_{j\in K}\left(Z_{n_k,j}(\hat{\theta}, P)\right)^2_+ > \delta\right\} \not\to 0.$$

Since $f_K : \mathbf{R}^m \to \mathbf{R}$ given by $f_K(x) = \sum_{j\in K}(x_j)^2_+$ is a continuous function, it is uniformly continuous on a compact set. Hence, for any $M > 0$ there exists a $\omega = \omega(M) > 0$ for which the left-hand side of (43) is bounded from above by

$$(44) \qquad P\left\{\sup_{\theta\in\Theta_0(P)}\max_{j\in K}|Z_{n_k,j}(\theta, P)| > M\right\}$$

$$+ P\left\{\max_{j\in K}|Z_{n_k,j}(\hat{\theta}_{n_k}, P) - Z_{n_k,j}(\hat{\theta}, P)| > \omega\right\}.$$

But it follows from assumption (i) that $Z_n(\theta, P)$ is asymptotically equicontinuous in the sense that for every $1 \le j \le m$ and $\delta > 0$,

$$\lim_{\eta\to 0}\limsup_{n\to\infty} P\left\{\sup_{\rho_{P,j}(\theta,\theta')<\eta}|Z_{n,j}(\theta, P) - Z_{n,j}(\theta', P)| > \delta\right\} = 0.$$

It thus follows from (40) and assumption (iii) that (44) tends to zero as $M \to \infty$ and $k \to \infty$. This yields a contradiction to (43), so (41) is established.

The asymptotic behavior of (35) is therefore given by the asymptotic behavior of (37). By assumption (i) and the continuous mapping theorem, (37) tends in distribution to

$$(45) \qquad \max_{K\subseteq\{1,\ldots,m\}}\sup_{\theta\in\Theta_0(K,P)}\sum_{j\in K}(Z_j(\theta, P))^2_+,$$

where $Z(\theta, P) = (Z_1(\theta, P), \ldots, Z_m(\theta, P))$ is a mean-zero multivariate Gaussian process with covariance kernel

$$\mathrm{Cov}(Z_j(\theta, P), Z_{j'}(\theta', P)) = E_P[g_j(X_i, \theta)g_{j'}(X_i, \theta')].$$

It remains to show that (45) is continuous at its $1 - \alpha$ quantile for appropriate values of α. To analyze this question, first note that (45) is a convex function

of $Z(\theta, P)$. It therefore follows from Theorem 11.1 of Davydov, Lifshits, and Smorodina (1998) that the distribution of (45) is continuous everywhere except possibly at zero. Next, note that

$$P\left\{\max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_0(K,P)} \sum_{j \in K} (Z_j(\theta, P))_+^2 \leq 0\right\} \leq P\left\{(Z_{j^*}(\theta^*, P))_+^2 \leq 0\right\} \leq \frac{1}{2},$$

where j^* and θ^* are as in assumption (iii). Hence, (45) is continuous at its $1 - \alpha$ quantile for all $\alpha < \frac{1}{2}$.

A.3. *Technical Details for Remark 2.7*

Let (X_i, Y_i), $i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P on $\mathbf{R}^2$. Let $\mu_X(P)$ denote the mean of the first component of the distribution P and let $\mu_Y(P)$ denote the mean of the second component of the distribution P. The parameter of interest, θ_0, is known to satisfy $\mu_X(P) \leq \theta_0 \leq \mu_Y(P)$. The identified set is therefore given by $\Theta_0(P) = \{\theta \in \mathbf{R} : \mu_X(P) \leq \theta \leq \mu_Y(P)\}$. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = (\mu_X(P) - \theta)_+^2 + (\theta - \mu_Y(P))_+^2.$$

The sample analog of $Q(\theta, P)$ is given by $\hat{Q}_n(\theta) = (\bar{X}_n - \theta)_+^2 + (\theta - \bar{Y}_n)_+^2$.

Let $a_n = n$ and suppose P is such that $E[|(X_i, Y_i)|^4] < \infty$ and that Cramer's condition holds, that is,

$$\limsup_{|s| \to \infty} |\psi_P(s)| < 1,$$

where $\psi_P(s)$ denotes the characteristic function of P. Assume further that $X_i \leq Y_i$ with probability 1 under P. This assumption is not essential, but it simplifies the analysis, while still allowing us to make our comparison. See Remark A.1 below for further discussion.

Let

$$\hat{\Theta}_{0,n} = \{\theta \in \mathbf{R} : n(\bar{X}_n - \theta)_+^2 + n(\theta - \bar{Y}_n)_+^2 \leq \lambda_n\},$$

where $\lambda_n > 0$ is an increasing sequence tending to infinity, but so slowly that $\lambda_n/n \to 0$. Chernozhukov, Hong, and Tamer (2007) suggested, for example, $\lambda_n = \log(n)$. Consider the confidence region given by

$$C_n' = \left\{\theta \in \mathbf{R} : n(\bar{X}_n - \theta)_+ + n(\theta - \bar{Y}_n)_+^2 \leq \hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)\right\}.$$

To obtain a second-order accurate expression for $\hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)$, first note that

$$(46) \qquad \hat{\Theta}_{0,n} = \left\{\theta \in \mathbf{R} : \bar{X}_n - \sqrt{\frac{\lambda_n}{n}} \leq \theta \leq \bar{Y}_n + \sqrt{\frac{\lambda_n}{n}}\right\}.$$

We therefore have that $\hat{r}_n(\hat{\Theta}_{0,n}, 1-\alpha)$ is the $1-\alpha$ quantile of

$$L_n(x) = \frac{1}{N_n} \sum_{1 \le i \le N_n} I\left\{ \max\left\{ b\left(\bar{X}_{n,b,i} - \bar{X}_n + \sqrt{\frac{\lambda_n}{n}}\right)_+^2, \right.\right.$$

$$\left.\left. b\left(\bar{Y}_n - \bar{Y}_{n,b,i} + \sqrt{\frac{\lambda_n}{n}}\right)_+^2 \right\} \le x \right\},$$

which, for $x \ge 0$, we may rewrite as

$$\frac{1}{N_n} \sum_{1 \le i \le N_n} I\left\{ \sqrt{b}(\bar{X}_{n,b,i} - \bar{X}_n) \le \sqrt{x} - \sqrt{\frac{\lambda_n b}{n}}, \right.$$

$$\left. \sqrt{b}(\bar{Y}_n - \bar{Y}_{n,b,i}) \le \sqrt{x} - \sqrt{\frac{\lambda_n b}{n}} \right\}.$$

Now consider

$$\tilde{L}_n(x,y) = \frac{1}{N_n} \sum_{1 \le i \le N_n} I\left\{ \sqrt{b}(\bar{X}_{n,b,i} - \bar{X}_n) \le x, \sqrt{b}(\bar{Y}_n - \bar{Y}_{n,b,i}) \le y \right\}$$

and the related

$$U_n(x,y) = \frac{1}{N_n} \sum_{1 \le i \le N_n} I\left\{ \sqrt{b}(\bar{X}_{n,b,i} - \mu_X(P)) \le x, \right.$$

$$\left. \sqrt{b}(\mu_Y(P) - \bar{Y}_{n,b,i}) \le y \right\}.$$

It follows from Lemma A.1 that

$$U_n(x,y) = J_b(x,y,P) + O_P\left(\sqrt{\frac{\log(k_n)}{k_n}}\right)$$

uniformly in x and y, where $J_b(x,y,P) = P\{\sqrt{b}(\bar{X}_n - \mu_X(P)) \le x, \sqrt{b} \times (\mu_Y(P) - \bar{Y}_n) \le y\}$. To see this, simply take $\delta = 4\sqrt{5}\sqrt{\log(k_n)/k_n}$ in equation (30) of Lemma A.1. From Lemma 5.4 of Hall (1992), we have, under our above assumptions on P, that

$$J_b(x,y,P) = \Phi_{\sigma_X(P),\sigma_Y(P),\rho(P)}(x,y) + \frac{1}{\sqrt{b}} f(x,y,P) + O_P\left(\frac{1}{b}\right)$$

uniformly in x and y, where $\Phi_{\sigma_X(P),\sigma_Y(P),\rho(P)}(x,y)$ is the bivariate normal cumulative distribution function with variances $\sigma_X^2(P)$ and $\sigma_Y^2(P)$ and covariance $\rho(P)$, and $f(x,y,P)$ is a smooth function of x and y that depends on P

through its second- and third-order cumulants. Now, since $\sqrt{b/n}$ is of smaller order than $\sqrt{\log(k_n)/k_n}$,

$$\sqrt{b}|\bar{X}_n - \mu_X(P)| = O_P(\sqrt{b/n}) \leq O_P(\sqrt{\log(k_n)/k_n}),$$

and similarly for $\bar{Y}_n$. Using these facts, we can argue as in the proof of Lemma 2 of Bertail (1997) that

$$\tilde{L}_n(x, y) = U_n(x, y) = O_P(\sqrt{\log(k_n)/k_n}).$$

Therefore,

$$\tilde{L}_n(x, y) = \Phi_{\sigma_X(P),\sigma_Y(P),\rho(P)}(x, y) + \frac{1}{\sqrt{b}}f(x, y, P)$$

$$+ O_P\left(\frac{1}{b} \vee \sqrt{\frac{\log(k_n)}{k_n}}\right)$$

uniformly in x and y. From this expression for $\tilde{L}_n(x, y)$, we can deduce that $\hat{c}_n = \inf\{x \in \mathbf{R} : \tilde{L}_n(x, x) \geq 1 - \alpha\}$ satisfies

$$\hat{c}_n = c + \frac{\delta}{\sqrt{b}} + O_P\left(\frac{1}{b} \vee \sqrt{\frac{\log(k_n)}{k_n}}\right),$$

where c is such that $\Phi_{\sigma_X(P),\sigma_Y(P),\rho(P)}(c, c) = 1 - \alpha$ and

$$\delta = -\frac{f(c, c)}{\nabla_{x,y}\Phi_{\sigma_X(P),\sigma_Y(P),\rho(P)}(x, y)|'_{x=c,y=c}(1, 1)}.$$

Hence,

$$(47) \qquad \hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha) = \left(\hat{c}_n + \sqrt{\frac{\lambda_n b}{n}}\right)^2$$

$$(48) \qquad = \left(c + \frac{\delta}{\sqrt{b}} + \sqrt{\frac{\lambda_n b}{n}} + O_P\left(\frac{1}{b} \vee \sqrt{\frac{\log k_n}{k_n}}\right)\right)^2.$$

Now consider the confidence region $\mathcal{C}_n$ given by Algorithm 2.1. First note that

$$\sup_{\theta \in \hat{\Theta}_{0,n}} a_n \hat{Q}_n(\theta) = \lambda_n.$$

From the above expression for $\hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)$, we therefore have that

$$P\left\{ \sup_{\theta \in \hat{\Theta}_{0,n}} a_n \hat{Q}_n(\theta) > \hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha) \right\} \to 1.$$

It follows that $\mathcal{C}_n$ is no larger than

$$\left\{ \theta \in \mathbf{R} : n(\bar{X}_n - \theta)_+^2 + n(\theta - \bar{Y}_n)_+^2 \le \hat{r}_n(\mathcal{C}_n', 1 - \alpha) \right\}$$

with probability tending to 1. From the above analysis, we have immediately that a second-order accurate expression for $\hat{r}_n(\mathcal{C}_n', 1 - \alpha)$ is

$$(49) \qquad \left(c + \frac{\delta}{\sqrt{b}} + \sqrt{\frac{b}{n}}\left(c + \frac{\delta}{\sqrt{b}} + \sqrt{\frac{\lambda_n b}{n}} \right) + O_P\left(\frac{1}{b} \vee \sqrt{\frac{\log(k_n)}{k_n}} \right) \right)^2.$$

Since $\sqrt{\frac{b}{n}}(c + \frac{\delta}{\sqrt{b}} + \sqrt{\lambda_n b/n})$ is of smaller order than $\sqrt{\lambda_n b/n}$, we expect that $\hat{r}_n(\mathcal{C}_n', 1 - \alpha)$ will be smaller to second order than $\hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)$. To illustrate this, consider the special case in which $\lambda_n = \log(n)$, as suggested by Chernozhukov, Hong, and Tamer (2007), and $b = n^{1/3}$, as suggested by Politis, Romano, and Wolf (1999). In this case, (48) simplifies to

$$c + \frac{\delta}{\sqrt{b}} + O_P\left(\sqrt{\frac{\log(n)}{n^{2/3}}} \right),$$

whereas (49) simplifies to

$$c + \frac{\delta}{\sqrt{b}} + O_P\left(\sqrt{\frac{\log(n^{2/3})}{n^{2/3}}} \right).$$

Hence, $\hat{r}_n(\mathcal{C}_n', 1 - \alpha)$ is smaller to second order than $\hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)$, as expected.

REMARK A.1: The assumption that $X_i \le Y_i$ with probability 1 under P was only used in the second-order comparison of Remark 2.7 to express $\hat{\Theta}_{0,n}$ as in (46). If we were to change $\hat{Q}_n(\theta)$, then this assumption can be removed as well. For example, if we consider

$$\hat{Q}_n(\theta) = \max\{n(\bar{X}_n - \theta)_+^2, n(\theta - \bar{Y}_n)_+^2\},$$

then we can express $\hat{\Theta}_{0,n}$ as in (46) regardless of whether $X_i \le Y_i$ with probability 1 under P or not. The rest of the analysis can be followed without any changes to illustrate the second-order benefits of the iterative approach.

A.4. *Technical Details for Example 2.3*

We begin with some preliminaries. We can write

$$\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) = \sup_{\theta \in \Theta_0(P)} \sum_{1 \leq j \leq m} \left(Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)]\right)_+^2,$$

where

$$Z_{n,j}(\theta, P) = \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} \left(g_j(X_i, \theta) - E_P[g_j(X_i, \theta)]\right).$$

Let $0 > \lambda_n \to 0$, but in a way that $\sqrt{b}\lambda_n \to 0$ and $\sqrt{n}\lambda_n \to -\infty$. For $K \subseteq \{1, \ldots, m\}$, define

$$\Theta'_n(K, P) = \left\{\theta \in \Theta_0(P) : \lambda_n < E_P[g_j(X_i, \theta)] \iff j \in K\right\},$$

$$\Theta_n(K, P) = \left\{\theta \in \Theta_0(P) : \lambda_n < E_P[g_j(X_i, \theta)] \text{ for all } j \in K\right\},$$

$$\Theta_0(K, P) = \left\{\theta \in \Theta_0(P) : E_P[g_j(X_i, \theta)] = 0 \text{ for all } j \in K\right\}.$$

Note that

$$\Theta_0(P) = \bigcup \{\Theta_n(K, P) : K \subseteq \{1, \ldots, m\}\}$$

$$= \bigcup \{\Theta'_n(K, P) : K \subseteq \{1, \ldots, m\}\}.$$

Below we will adopt the convention that the sum over the empty set and the supremum over the empty set are zero.

Next, note that for any sequence $P_n \in \mathbf{P}$, we have that

(50) $$\sup_{\theta \in \Theta_0(P_n)} a_b \hat{Q}_b(\theta)$$

$$= \max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{1 \leq j \leq m} \left(Z_{b,j}(\theta, P_n) + \sqrt{b} E_{P_n}[g_j(X_i, \theta)]\right)_+^2$$

$$\geq \max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} \left(Z_{b,j}(\theta, P_n) + \sqrt{b} E_{P_n}[g_j(X_i, \theta)]\right)_+^2$$

$$\geq \max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} \left(Z_{b,j}(\theta, P_n) + \sqrt{b}\lambda_n\right)_+^2$$

$$= \max_{K \subseteq \{1, \ldots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} \left(Z_{b,j}(\theta, P_n)\right)_+^2 + \Delta_{1,n}(P_n),$$

where

$$\Delta_{1,n}(P_n) = \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta_n(K,P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n) + \sqrt{b}\lambda_n)_+^2$$

$$- \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta_n(K,P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n))_+^2.$$

Moreover, $\Delta_{1,n}(P_n) = o_{P_n}(1)$. To see this, it suffices to show that for any $K \subseteq \{1, \dots, m\}$,

$$(51) \qquad \sup_{\theta \in \Theta_n(K,P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n))_+^2$$

$$- \sup_{\theta \in \Theta_n(K,P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n) + \sqrt{b}\lambda_n)_+^2 = o_{P_n}(1).$$

Since the left-hand side of (51) is bounded from below by

$$\sup_{\theta \in \Theta_n(K,P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n))_+^2 - (Z_{b,j}(\theta, P_n) + \sqrt{b}\lambda_n)_+^2 \geq 0,$$

it suffices to show that (51) is bounded from above by zero in probability. To this end, let $\hat{\theta}_n \in \Theta_n(K, P_n)$ be such that

$$\sup_{\theta \in \Theta_n(K,P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n))_+^2 \leq \sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n))_+^2 + \varepsilon_n$$

for $0 < \varepsilon_n \to 0$. Thus, (51) is bounded from above by

$$\sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n))_+^2 + \varepsilon_n - \sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n) + \sqrt{b}\lambda_n)_+^2.$$

Since the function $f_K : \mathbf{R}^m \to \mathbf{R}$ defined by $f_K(x) = \sum_{j \in K} (x_j)_+^2$ is continuous, it is uniformly continuous on a compact set. Hence, for any $\delta > 0$ and $M > 0$, there exists $\omega = \omega(M) > 0$ such that

$$(52) \qquad P_n \left\{ \sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n))_+^2 - \sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n) + \sqrt{b}\lambda_n)_+^2 > \delta \right\}$$

$$\leq P_n \left\{ \max_{1 \leq j \leq m} |Z_{b,j}(\hat{\theta}_n, P_n)| > M \right\} + P_n \{ \sqrt{b}\lambda_n < -\omega \}$$

$$\leq P_n \left\{ \sup_{\theta \in \Theta} \max_{1 \leq j \leq m} |Z_{b,j}(\theta, P_n)| > M \right\} + I\{ \sqrt{b}\lambda_n < -\omega \}.$$

By assumption (i),

$$\sup_{\theta \in \Theta} \max_{1 \le j \le m} |Z_{b,j}(\theta, P_n)| = O_{P_n}(1).$$

It therefore follows from the assumption that $\sqrt{b}\lambda_n \to 0$ that (52) tends to zero as $n \to \infty$ and then $M \to \infty$. The desired claim thus follows.

Similarly, we have for any sequence $P_n \in \mathbf{P}$ that

(53)
$$\sup_{\theta \in \Theta_0(P_n)} a_n \hat{Q}_n(\theta)$$

$$= \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta'_n(K,P_n)} \sum_{1 \le j \le m} \left(Z_{n,j}(\theta, P_n) + \sqrt{n} E_{P_n}[g_j(X_i, \theta)]\right)_+^2$$

$$\le \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta'_n(K,P_n)} \sum_{j \in K} (Z_{n,j}(\theta, P_n))_+^2$$

$$+ \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta'_n(K,P_n)} \sum_{j \notin K} \left(Z_{n,j}(\theta, P_n) + \sqrt{n} E_{P_n}[g_j(X_i, \theta)]\right)_+^2$$

$$\le \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta_n(K,P_n)} \sum_{j \in K} (Z_{n,j}(\theta, P_n))_+^2 + \Delta_{2,n}(P_n),$$

where

$$\Delta_{2,n}(P_n) = \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta'_n(K,P_n)} \sum_{j \notin K} \left(Z_{n,j}(\theta, P_n) + \sqrt{n} E_{P_n}[g_j(X_i, \theta)]\right)_+^2.$$

Moreover, $\Delta_{2,n}(P_n) = o_{P_n}(1)$. To see this, note that

$$\sup_{\theta \in \Theta'_n(K,P_n)} \sum_{j \notin K} \left(Z_{n,j}(\theta, P_n) + \sqrt{n} E_{P_n}[g_j(X_i, \theta)]\right)_+^2$$

$$\le m \left(\sup_{\theta \in \Theta} \max_{1 \le j \le m} |Z_{n,j}(\theta, P_n)| + \sqrt{n}\lambda_n\right)_+^2.$$

By assumption (i),

$$\sup_{\theta \in \Theta} \max_{1 \le j \le m} |Z_{n,j}(\theta, P_n)| = O_{P_n}(1).$$

The desired claim thus follows from the assumption that $\sqrt{n}\lambda_n \to -\infty$. In fact, we have that $\Delta_{2,n}(P_n)$ is identically equal to zero with probability tending to 1.

We now use these facts to argue by contradiction that the required condition (21) holds. If the result were false, then there would exist a subsequence n_k and a corresponding sequence $P_{n_k} \in \mathbf{P}$ such that

$$\sup_{x \in \mathbf{R}} \left\{J_{b_{n_k}}(x, P_{n_k}) - J_{n_k}(x, P_{n_k})\right\} \to \delta$$

for some $\delta > 0$. It follows from (50) and (53) that

$$(54) \qquad \sup_{x \in \mathbf{R}} \{ \tilde{J}_{b_{n_k}}(x, P_{n_k}) - \bar{J}_{n_k}(x, P_{n_k}) \} \not\to 0,$$

where

$$\tilde{J}_{b_{n_k}}(x, P_{n_k}) = P_{n_k} \left\{ \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_{n_k}(K, P_{n_k})} \sum_{j \in K} \left(Z_{b_{n_k}, j}(\theta, P_{n_k}) \right)_+^2 \right.$$

$$\left. + \Delta_{1, n_k}(P_{n_k}) \le x \right\},$$

$$\bar{J}_{n_k}(x, P_{n_k}) = P_{n_k} \left\{ \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_{n_k}(K, P_{n_k})} \sum_{j \in K} \left(Z_{n_k, j}(\theta, P_{n_k}) \right)_+^2 \right.$$

$$\left. + \Delta_{2, n_k}(P_{n_k}) \le x \right\}.$$

By assumption (ii) and Theorem 3.85 of Aliprantis and Border (2006), the set of all nonempty closed subsets of Θ is a compact metric space with respect to the Hausdorff metric

$$\bar{\rho}_H(A, B) = \inf\{ \eta > 0 : A \subseteq B^\eta, B \subseteq A^\eta \},$$

where

$$A^\eta = \bigcup \{ a' \in \Theta : \bar{\rho}(a', a) < \eta \text{ for some } a \in A \}.$$

Hence, for each $K \subseteq \{1, \ldots, m\}$ for which $\Theta_{n_k}(K, P_{n_k})$ is nonempty infinitely often, there exists $\emptyset \ne \Theta^*(K) \subseteq \Theta$ such that $\Theta_{n_k}(K, P_{n_k})$ converges to $\Theta^*(K)$ under $\bar{\rho}_H$ along a subsequence. For any $K \subseteq \{1, \ldots, m\}$ for which $\Theta_{n_k}(K, P_{n_k})$ is nonempty only finitely often, let $\Theta^*(K) = \emptyset$. Note that by assumption (iii) and the pigeonhole principle there exists $1 \le j^* \le m$ such that infinitely often there exists $\theta_{n_k}^* \in \Theta_{n_k}(\{j^*\}, P_{n_k})$ such that $\mathrm{Var}_{P_{n_k}}[g_{j^*}(X_i, \theta_{n_k}^*)] \ge \varepsilon$.

By assumption (i), we have that

$$\sup_{\theta \in \Theta} \max_{1 \le j \le m} \left| Z_{b_{n_k}, j}(\theta, P_{n_k}) \right| = O_{P_{n_k}}(1),$$

$$\sup_{\theta \in \Theta} \max_{1 \le j \le m} \left| Z_{n_k, j}(\theta, P_{n_k}) \right| = O_{P_{n_k}}(1).$$

Hence, along a subsequence, both $Z_{b_{n_k}}(\theta, P_{n_k})$ and $Z_{n_k}(\theta, P_{n_k})$ converge to mean-zero multivariate Gaussian processes. Since the covariance kernel of these limiting processes is given by the limit of the covariance kernels of $Z_{b_{n_k}}(\theta, P_{n_k})$ and $Z_{n_k}(\theta, P_{n_k})$, which are identical, the covariance kernels of

the limiting processes must also be identical. The distributions of the limiting processes must therefore be identical.

In summary, there exists a subsequence n_{k_ℓ} along which $\rho_H(\Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}}), \Theta^*(K)) \to 0$ for all $K \subseteq \{1, \ldots, m\}$, and $Z_{b_{n_{k_\ell}}}(\theta, P_{n_{k_\ell}})$ and $Z_{n_{k_\ell}}(\theta, P_{n_{k_\ell}})$ both converge in distribution $Z^*(\theta)$, a mean-zero multivariate Gaussian process. Let $J^*(x)$ denote the distribution function of

$$(55) \qquad \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta^*(K)} \sum_{j \in K} (Z_j^*(\theta))_+^2.$$

Furthermore, there exist $1 \leq j^* \leq m$ and $\theta^* \in \Theta^*(\{j^*\})$ such that $Z_{j^*}^*(\theta^*)$ is not degenerate. Since (55) is a convex function of $Z^*(\theta)$, it follows from Theorem 11.1 of Davydov, Lifshits, and Smorodina (1998) that $J^*(x)$ is continuous on $(0, \infty)$.

To complete the argument, we argue that

$$\sup_{x \in \mathbf{R}} \left| \tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}}) - J^*(x) \right| \to 0,$$

$$\sup_{x \in \mathbf{R}} \left| \bar{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}}) - J^*(x) \right| \to 0,$$

which will lead to a contradiction of (54). Note that the distribution functions are identically equal to zero for $x < 0$, so it suffices to show that

$$\sup_{x \geq 0} \left| \tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}}) - J^*(x) \right| \to 0,$$

$$\sup_{x \geq 0} \left| \bar{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}}) - J^*(x) \right| \to 0.$$

For this purpose, we will use Lemma A.2. To apply this lemma, we first argue that

$$(56) \qquad \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} \left(Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}}) \right)_+^2$$

$$- \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta^*(K)} \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}}))_+^2$$

and

$$(57) \qquad \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} \left(Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}}) \right)_+^2$$

$$- \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta^*(K)} \sum_{j \in K} (Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}}))_+^2$$

both converge in probability to zero. We only establish the result for (56), as essentially the same argument will establish the result for (57). It suffices to show that

$$(58) \quad \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} \left(Z_{b_{n_{k_\ell}}, j}\left(\theta, P_{n_{k_\ell}}\right) \right)^2_+$$

$$- \sup_{\theta \in \Theta^*(K)} \sum_{j \in K} \left(Z_{b_{n_{k_\ell}}, j}\left(\theta, P_{n_{k_\ell}}\right) \right)^2_+ = o_{P_{n_{k_\ell}}}(1).$$

We first argue that the left-hand side of (58) is bounded above by zero in probability. To this end, let $\hat{\theta}_\ell \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})$ be such that

$$\sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} \left(Z_{b_{n_{k_\ell}}, j}\left(\theta, P_{n_{k_\ell}}\right) \right)^2_+ \leq \sum_{j \in K} \left(Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}_\ell, P_{n_{k_\ell}}\right) \right)^2_+ + \varepsilon_\ell$$

for $0 < \varepsilon_\ell \to 0$, and let $\hat{\theta}^*_\ell \in \Theta^*(K)$ be such that $\bar{\rho}(\hat{\theta}_\ell, \hat{\theta}^*_\ell) \to 0$. With $\hat{\theta}_\ell$ and $\hat{\theta}^*_\ell$ so defined, the left-hand side of (58) is bounded above by

$$\sum_{j \in K} \left(Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}_\ell, P_{n_{k_\ell}}\right) \right)^2_+ + \varepsilon_\ell - \sum_{j \in K} \left(Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}^*_\ell, P_{n_{k_\ell}}\right) \right)^2_+.$$

Since the function $f_K : \mathbf{R}^m \to \mathbf{R}$ defined by $f_K(x) = \sum_{j \in K}(x_j)^2_+$ is continuous, it is uniformly continuous on a compact set. Hence, for any $\delta > 0$ and $M > 0$, there exists $\omega = \omega(M) > 0$ such that

$$(59) \quad P_{n_{k_\ell}} \left\{ \sum_{j \in K} \left(Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}_\ell, P_{n_{k_\ell}}\right) \right)^2_+ - \sum_{j \in K} \left(Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}^*_\ell, P_{n_{k_\ell}}\right) \right)^2_+ > \delta \right\}$$

$$\leq P_{n_{k_\ell}} \left\{ \max_{1 \leq j \leq m} \left| Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}_\ell, P_{n_{k_\ell}}\right) \right| > M \right\}$$

$$+ P_{n_{k_\ell}} \left\{ \max_{j \in K} \left| Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}_\ell, P_{n_{k_\ell}}\right) - Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}^*_\ell, P_{n_{k_\ell}}\right) \right| > \omega \right\}$$

$$\leq P_{n_{k_\ell}} \left\{ \sup_{\theta \in \Theta} \max_{1 \leq j \leq m} \left| Z_{b_{n_{k_\ell}}, j}\left(\theta, P_{n_{k_\ell}}\right) \right| > M \right\}$$

$$+ P_{n_{k_\ell}} \left\{ \max_{1 \leq j \leq m} \left| Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}_\ell, P_{n_{k_\ell}}\right) - Z_{b_{n_{k_\ell}}, j}\left(\hat{\theta}^*_\ell, P_{n_{k_\ell}}\right) \right| > \omega \right\}.$$

By assumption (i),

$$(60) \quad \sup_{\theta \in \Theta} \max_{1 \leq j \leq m} \left| Z_{b_{n_{k_\ell}}, j}\left(\theta, P_{n_{k_\ell}}\right) \right| = O_{P_{n_{k_\ell}}}(1).$$

Assumption (i) also implies that for any $\omega > 0$ and all $1 \le j \le m$,

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}} P \left\{ \sup_{\rho_{P,j}(\theta,\theta') < \eta} |Z_{n,j}(\theta, P) - Z_{n,j}(\theta', P)| > \omega \right\} = 0.$$

It therefore follows from the assumption that $\bar{\rho}(\hat{\theta}_\ell, \hat{\theta}_\ell^*) \to 0$ that (59) tends to zero as $\ell \to \infty$ and then $M \to \infty$. By interchanging the roles of $\Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})$ and $\Theta^*(K)$, we see that the left-hand side of (58) is also bounded below by zero in probability. The desired result follows.

Since $\Delta_{1,n_{k_\ell}}(P_{n_{k_\ell}})$ and $\Delta_{2,n_{k_\ell}}(P_{n_{k_\ell}})$ both converge in probability to zero, we have by the continuous mapping theorem that $\tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}})$ and $\bar{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}})$ both converge in distribution to $J^*(x)$. Since $J^*(x)$ is continuous on $(0, \infty)$, we have for every $x > 0$ that

$$\tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}}) \to J^*(x),$$

$$\bar{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}}) \to J^*(x).$$

To apply Lemma A.2, we therefore need only show that these convergences hold at $x = 0$.

Consider first $\tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}})$. Since $\Delta_{1,n_{k_\ell}}(P_{n_{k_\ell}}) \le 0$, we have that

$$\tilde{J}_{b_{n_{k_\ell}}}(0, P_{n_{k_\ell}}) \ge P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} \left(Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}) \right)_+^2 \le 0 \right\}$$

$$= P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}) \le 0 \right\}.$$

Conversely, since $\Delta_{1,n_{k_\ell}}(P_{n_{k_\ell}}) = o_{P_{n_{k_\ell}}}(1)$, we have for any $\varepsilon > 0$ that

$$\limsup_{\ell \to \infty} \tilde{J}_{b_{n_{k_\ell}}}(0, P_{n_{k_\ell}})$$

$$\le \limsup_{\ell \to \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} \left(Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}) \right)_+^2 \le \varepsilon \right\}$$

$$\le \limsup_{\ell \to \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} \left(Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}) \right)_+^2 \le \varepsilon \right\}$$

$$= \limsup_{\ell \to \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1,\ldots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}) \le \sqrt{\varepsilon} \right\}.$$

But arguing as was done to establish that (56) and (57) both tended to zero in probability, we have by the continuous mapping theorem that

$$(61) \qquad \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}}) \xrightarrow{\mathcal{L}} \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta^*(K)} \max_{j \in K} Z_j^*(\theta).$$

By letting $\varepsilon \to 0$ and noting that $J^*(0)$ is equal to the probability that (61) is less than or equal to zero, the desired result follows.

Now consider $\bar{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}})$. Since $\Delta_{2,n_{k_\ell}}(P_{n_{k_\ell}})$ is identically equal to zero with probability tending to 1, we have that

$$\lim_{\ell \to \infty} \bar{J}_{n_{k_\ell}}(0, P_{n_{k_\ell}})$$

$$\leq \lim_{\ell \to \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} (Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}}))_+^2 \leq 0 \right\}$$

$$= \lim_{\ell \to \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}}) \leq 0 \right\}.$$

But arguing as was done to establish that (56) and (57) both tended to zero in probability, we have by the continuous mapping theorem that

$$\max_{K \subseteq \{1,\dots,m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}})$$

converges in distribution to the right-hand side of (61). To complete the proof, note that $J^*(0)$ is equal to the probability that the right-hand side of (61) is less than or equal to zero.

REFERENCES

ALIPRANTIS, C. D., AND K. C. BORDER (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*. New York: Springer. [206]

ANDREWS, D. W. K. (2000): "Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space," *Econometrica*, 68, 399–405. [180]

BAHADUR, R. R., AND L. J. SAVAGE (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *Annals of Mathematical Statistics*, 27, 1115–1122. [170]

BERTAIL, P. (1997): "Second-Order Properties of an Extrapolated Bootstrap Without Replacement Under Weak Assumptions," *Bernoulli*, 3, 149–179. [201]

BUGNI, F. (2007): "Bootstrap Inference in Partially Identified Models," Working Paper. [180]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2004): "Inference for Identified Parameter Sets in Econometric Models," Working Paper, Department of Economics, MIT. [186]

——— (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [170,180,181,199,202]

DAVYDOV, Y. A., M. A. LIFSHITS, AND N. V. SMORODINA (1998): *Local Properties of Distributions of Stochastic Functionals*. Providence, RI: American Mathematical Society. [179,199,207]

HALL, P. (1992): *The Bootstrap and Edgeworth Expansion*. New York: Springer. [200]

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [171]

KATZ, L. F., AND D. H. AUTOR (1999): "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics*, Vol. 3A, ed. by O. Ashenfelter and D. Card. Amsterdam: North-Holland, 1463–1555. [186]

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses* (Third Ed.). New York: Springer. [185]

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions With Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546. [178]

POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer. [173,176, 177,187,202]

POLLARD, D. (1984): *Convergence of Stochastic Proceses*. New York: Springer. [195]

ROMANO, J. P. (2004): "On Non-Parametric Testing, the Uniform Behavior of the *t*-Test, and Related Problems," *Scandinavian Journal of Statistics*, 31, 567–584. [171]

ROMANO, J. P., AND A. M. SHAIKH (2006): "Inference for the Identified Set in Partially Identified Econometric Models," Technical Report 2006-10, Department of Statistics, Stanford University. [180]

——— (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference—Special Issue in Honor of Ted Anderson*, 138, 2786–2807. [170,171,182,187]

——— (2010): "Supplement to 'Inference for the Identified Set in Partially Identified Econometric Models'," *Econometrica Supplemental Material*, 78, http://www.econometricsociety.org/ecta/Supmat/6706_data and programs.zip. [172]

ROMANO, J. P., AND M. WOLF (2005): "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing," *Journal of the American Statistical Association*, 100, 94–108. [175]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer. [178,183]

Depts. of Economics and Statistics, Sequoia Hall, Stanford University, Stanford, CA 94305-4065, U.S.A.; romano@stanford.edu

and

Dept. of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, U.S.A.; amshaikh@uchicago.edu.