Inference for Identifiable Parameters in Partially Identified Econometric Models

Joseph P. Romano
Department of Statistics
Stanford University
romano@stat.stanford.edu

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

September 2006

Abstract

This paper considers the problem of inference for partially identified econometric models. The class of models studied are defined by a population objective function $Q(\theta, P)$ for $\theta \in \Theta$. The second argument indicates the dependence of the objective function on P, the distribution of the observed data. Unlike the classical extremum estimation framework, 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 some unknown point $\theta \in \Theta_0(P)$, where $\Theta_0(P) = \arg\min_{\theta \in \Theta} Q(\theta, P)$, and so we seek random sets that contain each $\theta \in \Theta_0(P)$ with at least some prespecified probability asymptotically. We also consider situations where the object of interest is the image of some point $\theta \in \Theta_0(P)$ under a known function. Computationally intensive, yet feasible procedures for constructing random sets satisfying the desired coverage property under weak assumptions are provided. We also provide conditions under which the confidence regions are uniformly consistent in level. To do this, we first derive new uniformity results about subsampling that are of independent interest.

KEYWORDS: Extremum Estimation, Partially Identified Model, Incomplete Model, Identified Set, Identifiable Parameter, Subsampling, Uniform Coverage, Confidence Region, Interval Regression, Moment Inequalities

1 Introduction

This paper provides computationally intensive, yet feasible methods for inference for a large class of partially identified econometric models. A partially identified model is a model in which the parameter of interest is not uniquely defined by the distribution of the observed data. Such models arise naturally in many parts of empirical work in economics. The class of models considered is defined by a population objective function $Q(\theta, P)$ for $\theta \in \Theta$. The second argument indicates the dependence of the objective function on P, the distribution of the observed data. Unlike the classical extremum estimation framework, 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 first of these two goals. The object of interest is some unknown point $\theta \in \Theta_0(P)$, where

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

We henceforth refer to any $\theta \in \Theta_0(P)$ as an *identifiable parameter* and $\Theta_0(P)$ as the *identified set*. In this instance, 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. That is, for all $\theta \in \Theta_0(P)$ we have

$$\liminf_{n \to \infty} P\{\theta \in \mathcal{C}_n\} \geq 1 - \alpha \ . \tag{2}$$

We term such sets as *confidence regions for identifiable parameters 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 construct random sets $\mathcal{C}_n$ satisfying (2) under weak assumptions on P. These assumptions are formulated in terms of restrictions on the asymptotic behavior of the estimate of the population objective function, $\hat{Q}_n(\theta)$. Most often $\hat{Q}_n(\theta) = Q(\theta, \hat{P}_n)$ for some estimate $\hat{P}_n$ of P. Our construction is based on the usual duality between confidence regions and hypothesis testing: we invert tests of each of the null hypotheses $H_\theta : \theta \in \Theta_0(P)$ for $\theta \in \Theta$ that control the usual probability of a Type 1 error at level α. Typically, Θ is a subset of Euclidean space and so the number of null hypotheses may be uncountably infinite. The idea of inverting tests using computer-intensive methods can be traced to DiCiccio and Romano (1990).

In the second goal, the object of interest is the identified set, $\Theta_0(P)$, itself. In this case, given i.i.d. data $X_i, i = 1, \ldots, n$, generated from P, we seek instead 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. The problem of constructing such sets is treated in a companion paper Romano and Shaikh (2006).

Our results on confidence regions for identifiable parameters build upon the earlier work of Chernozhukov et al. (2004), who were the first to consider inference for the same class of partially identified models. We show that our construction produces smaller confidence regions than theirs without violating the desired coverage property. In fact, the confidence regions described in this paper may even be smaller in a first-order asymptotic sense.

Our analysis also extends the work of Imbens and Manski (2004), who analyze 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. The authors discuss different definitions for confidence regions in the context of this setting and argue forcefully that it may be desirable to demand validity of confidence regions not just for each fixed probability distribution P satisfying weak assumptions, 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, show 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) extends this non-existence result to a number of other problems. We provide restrictions on $\mathbf{P}$ under which both types of confidence regions suggested in this paper have good uniform behavior. Concretely, we provide conditions under which $\mathcal{C}_n$ satisfies

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

By analogy with our earlier terminology, sets satisfying (3) are referred to as *confidence regions for identifiable parameters 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,

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

In order to provide conditions under which these coverage properties hold, we first derive more general conditions under which confidence regions for a parameter constructed using subsampling are uniformly consistent in level. These results on subsampling are of interest independent of their role in constructing sets satisfying (3).

We have so far assumed that the object of interest is either an identifiable parameter, $\theta \in \Theta_0(P)$. More generally, the object of interest may be the image of an identifiable parameter 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,

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

We refer to a set $\mathcal{C}_n^f$ as a *confidence region for a function of identifiable parameters that is pointwise consistent in level* if for any $\lambda \in \Lambda_0(P)$ we have that

$$\liminf_{n\to\infty} P\{\lambda \in \mathcal{C}_n^f\} \geq 1-\alpha \ . \tag{5}$$

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

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

By analogy with our earlier terminology, sets satisfying (6) are referred to as *confidence regions for a function of identifiable parameters that are uniformly consistent in level.*

We adapt our constructions of confidence regions for identifiable parameters to provide constructions of confidence sets satisfying these alternative coverage requirements.

Methods for drawing inferences in partially identified models have allowed economists to solve many empirical problems that previously either were intractable or relied on untenable assumptions to achieve identification. Manski (2003) has argued forcefully against invoking such assumptions to make inferences in the context of missing outcome data and treatment response, as it degrades the credibility of inferences made under those assumptions. Another class of models that may only be partially identified is given by game-theoretic models with multiple equilibria. When confronted with such models, researchers often impose identification by assuming some sort of an ad hoc equilibrium selection mechanism. More recently, the empirical literature has explored inferences based on exploiting only those restrictions implied by equilibrium behavior that do not depend on the particular equilibrium being played by the agents. These restrictions are often defined by a system of moment inequalities, which is a special case of the class of models considered in this paper. Cilberto and Tamer (2004) and Borzekowski and Cohen (2005), for example, use this idea together with the techniques of Chernozhukov et al. (2004) for inference in partially identified models to analyze an entry model and a coordination game, respectively. Benkard et al. (2005) consider more generally the problem of making inferences in dynamic models of imperfect competition, where, in the absence of additional assumptions, the unknown parameters of the model are naturally restricted to a set defined by moment inequalities. Andrews et al. (2004) and Pakes et al. (2005) provide other economic applications of systems of moment inequalities, but develop independent methods of inference for systems of moment inequalities distinct from the methods of Chernozhukov et al. (2004). These methods are, however, more conservative than those of Chernozhukov et al. (2004), and thus more conservative than the methods developed in this paper as well.

The remainder of the paper is organized as follows. We begin in Section 2 by providing a number of concrete examples that fall within the class of models we are considering. These examples not only serve as motivation, but will also be useful later for illustrating our methodology. In Section 3, we analyze the problem of constructing confidence sets that satisfy the coverage requirements (2) and (3). In order to do so, we first develop some useful results about the coverage probability of confidence sets constructed using subsampling for a fixed distribution P and also uniformly over a class of distributions $\mathbf{P}$. We then extend this methodology to construct confidence regions satisfying (5) and (6). Finally, we conclude and provide directions for future research in Section 4.

2 Some Motivating Examples

In this section, we provide motivation for our study of partially identified models by describing several specific examples of such models. We will return to these examples repeatedly in the sequel as illustrations of our methodology.

Example 2.1 (*One-Sided Mean*) Perhaps the simplest example of a partially identified model is given by the following setup. Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P on $\mathbf{R}$. Denote by $\mu(P)$ the mean of the distribution P. The parameter of interest, θ_0, is known to satisfy $\theta_0 \geq \mu(P)$. For example, θ_0 might be the mean of an unobservable random variable Z_i with distribution Q on $\mathbf{R}$ that is known to satisfy $Z_i \geq X_i$, and so $\mu(Q) \geq \mu(P)$. The identified set is therefore given by $\Theta_0(P) = \{\theta \in \mathbf{R} : \theta \geq \mu(P)\}$. We may characterize this set as the set of minimizers of the population objective function

$$Q(\theta, P) = (\mu(P) - \theta)^2_+ \ ,$$

where the notation $(a)_+$ is used as shorthand for $\max\{a, 0\}$. The sample analog of $Q(\theta, P)$ is given by $\hat{Q}_n(\theta) = (\bar{X}_n - \theta)^2_+$. ■

Example 2.2 (*Two-Sided Mean*) A natural generalization of Example 2.1 is to consider bivariate random variables. To this end, 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 $\mu_Y(P)$ the mean of the second component of the distribution P. The parameter of interest, θ_0, is known to satisfy $\mu_X(P) \leq \theta_0 \leq \mu_Y(P)$. For example, θ_0 might be the mean of another distribution Q on $\mathbf{R}$ that is known to satisfy $\mu_X(P) \leq \mu(Q) \leq \mu_Y(P)$. 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 $\hat{Q}_n(\theta) = (\bar{X}_n - \theta)^2_+ + (\theta - \bar{Y}_n)^2_+$ ■

Remark 2.1 Let $W_i \in [0,1]$ and $D_i \in \{0,1\}$. For example, W_i may be the answer to the question, "Do you vote Republican or Democrat?", and D_i may be the indicator variable for whether the person asked chooses to answer the question. Suppose the researcher observes an i.i.d. sequence $(W_i D_i, D_i), i = 1, \ldots, n$, with distribution P; i.e., W_i is observed if and only if $D_i = 1$. The parameter of interest, $\theta_0 = E\{W_i\}$, is not determined by the distribution of the observed data, but the researcher can say with certainty that θ_0 satisfies $E_P\{W_i D_i\} \leq \theta_0 \leq E_P\{W_i D_i + 1 - D_i\}$. By identifying $X_i = W_i D_i$ and $Y_i = W_i D_i + 1 - D_i$, this example can be seen to be a special case of Example 2.2. This example is analyzed in detail by Imbens and Manski (2004). ■

Example 2.3 (*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 Q on $\mathbf{R}^k \times \mathbf{R} \times \mathbf{R} \times \mathbf{R}$. The parameter of interest, θ_0, is known to satisfy $E_Q\{Y_i^* | X_i\} = X_i' \theta_0$, 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 Q. 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

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

The sample analog $\hat{Q}_n(\theta)$ of $Q(\theta, P)$ is given by replacing expectations with appropriately defined estimators, the nature of which may depend on further assumptions about P. Manski and Tamer (2002) characterize the identified set in this setting and also consider the case where Y_i^* is observed, but X_i is interval-censored. ■

Remark 2.2 A Tobit-like model is a special case of the previous example 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 $Y_i^* \leq 0$. It is worthwhile to note that conventional approaches to inference for such a model, while enforcing identification, rely on the much stronger assumption that $\epsilon_i = Y_i^* - X_i'\theta_0$ is independent of X_i. By allowing for only partial identification, we will be able to draw inferences under much weaker assumptions. ■

Example 2.4 (*Moment Inequalities*) Consider the following generalization of Examples 2.1 and 2.2. 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 $\Theta_0(P) = \{\theta \in \mathbf{R}^l : E_P\{g_j(X_i, \theta)\} \leq 0 \ \forall j \text{ s.t. } 1 \leq j \leq m\}$. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = \sum_{1 \leq j \leq m} (E_P\{g_j(X_i, \theta)\})_+^2 ,$$

or equivalently as $\{\theta \in \Theta : Q(\theta, P) = 0\}$. The sample analog of $Q(\theta, P)$ is given by $\hat{Q}_n(\theta) = \sum_{1 \leq j \leq m}(\frac{1}{n}\sum_{1 \leq i \leq n} g_j(X_i, \theta))_+^2$. This choice of $Q(\theta, P)$ is especially noteworthy because it is used by Cilberto and Tamer (2004), Benkard et al. (2005) and Borzekowski and Cohen (2005) in their empirical applications. Since any equality restriction may be thought of as two inequality restrictions, this example may also be viewed as a generalization of the method of moments to allow for equality and inequality restrictions on the moments, rather than just equality restrictions. ■

Remark 2.3 A prominent example of an econometric model which gives rise to moment inequalities is an entry model. See, for example, Andrews et al. (2004) or Cilberto and Tamer (2004) for a detailed description of such models and a derivation of the inequalities. Briefly, consider an entry model with two firms and let X_i be the indicator for the event "firm 1 enters". Because of the multiplicity of Nash equilibria, the model only gives upper and lower *bounds*, $L(\theta)$ and $U(\theta)$, on the probability of this event as a function of the unknown parameter, θ, of the econometric model. It is therefore natural to use the functions

$$\begin{aligned} g_1(X_i, \theta) &= L(\theta) - X_i \\ g_2(X_i, \theta) &= X_i - U(\theta) \end{aligned}$$

as a basis for inference in such a model. ■

3 Confidence Regions for Identifiable Parameters

In this section, we consider the problem of constructing confidence regions for identifiable parameters. We begin in Section 3.1 with a general result about subsampling that will be used frequently throughout the remainder of the paper. For the sake of continuity, the proof of the result, which is lengthy, is deferred to the appendix. In Section 3.2, we then treat the construction of sets satisfying (2), before turning our attention in Section 3.3 to the problem of constructing sets satisfying (3). In Section 3.4, we generalize these constructions to confidence regions for functions of identifiable parameters.

3.1 A Useful Result about Subsampling

The following theorem describes conditions under which confidence sets constructed using subsampling are asymptotically valid for a fixed P as well as uniformly over $\mathbf{P}$. The conditions for a fixed P are distinct from those described in Politis et al. (1999) and can accomodate certain degenerate situations not covered by the results presented there. As we will see below, this feature of the conditions will be essential for the types of problems we are considering.

Theorem 3.1 *Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P. Let $\vartheta(P)$ be a real-valued parameter and let $\hat{\vartheta}_n$ be some estimator of $\vartheta(P)$. Denote by $J_n(\cdot, P)$ the distribution of the root $\tau_n(\hat{\vartheta}_n - \vartheta(P))$, where τ_n is a sequence of known constants. 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*

$$L_n(x) = \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\{\tau_b(\hat{\vartheta}_{n,b,i} - \vartheta(P)) \leq x\} \ , \tag{7}$$

where $\hat{\vartheta}_{n,b,i}$ denotes the estimate $\hat{\vartheta}_n$ evaluated at the ith subset of data of size b. Finally, let $L_n^{-1}(1-\alpha) = \inf\{x \in \mathbf{R} : L_n(x) \geq 1 - \alpha\}$. Then, for $\alpha \in (0,1)$, the following statements are true:

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

$$\liminf_{n\to\infty} P\{\tau_n(\hat{\vartheta}_n - \vartheta(P)) \leq L_n^{-1}(1-\alpha)\} \geq 1 - \alpha \ .$$

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

$$\liminf_{n\to\infty} P\{\tau_n(\hat{\vartheta}_n-\vartheta(P))\ge L_n^{-1}(\alpha)\}\ge 1-\alpha\ .$$

(iii) *If* $\lim_{n\to\infty}\sup_{x\in\mathbf{R}}|J_b(x,P)-J_n(x,P)|=0$, *then*

$$\liminf_{n\to\infty} P\{L_n^{-1}(\frac{\alpha}{2})\le\tau_n(\hat{\vartheta}_n-\vartheta(P))\le L_n^{-1}(1-\frac{\alpha}{2})\}\ge 1-\alpha\ .$$

Morever, we have that:

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

$$\liminf_{n\to\infty}\inf_{P\in\mathbf{P}} P\{\tau_n(\hat{\vartheta}_n-\vartheta(P))\le L_n^{-1}(1-\alpha)\}\ge 1-\alpha\ .$$

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

$$\liminf_{n\to\infty}\inf_{P\in\mathbf{P}} P\{\tau_n(\hat{\vartheta}_n-\vartheta(P))\ge L_n^{-1}(\alpha)\}\ge 1-\alpha\ .$$

(vi) *If* $\lim_{n\to\infty}\sup_{P\in\mathbf{P}}\sup_{x\in\mathbf{R}}|J_b(x,P)-J_n(x,P)|=0$, *then*

$$\liminf_{n\to\infty}\inf_{P\in\mathbf{P}} P\{L_n^{-1}(\frac{\alpha}{2})\le\tau_n(\hat{\vartheta}_n-\vartheta(P))\le L_n^{-1}(1-\frac{\alpha}{2})\}\ge 1-\alpha\ .$$

Proof: See appendix. ■

Remark 3.1 The subsampling distribution $L_n(x)$ defined by (7) differs from the typical definition given in Politis et al. (1999) in that it centers about the true value of the parameter $\vartheta(P)$ instead of its estimate $\hat{\vartheta}_n$. Strictly speaking, it is therefore not possible to compute $L_n(x)$ as it is defined above. We may, however, compute $L_n(x)$ under the null hypothesis that $\vartheta(P)$ equals a particular value. For our purposes below, this will suffice. ■

3.2 Pointwise Consistency in Level

We assume in the remainder of the paper that $\inf_{\theta\in\Theta} Q(\theta,P)=0$. Note that this restriction is without loss of generality, as it can always be enforced simply by subtracting the quantity $\inf_{\theta\in\Theta} Q(\theta,P)$ from the objective function whenever $\inf_{\theta\in\Theta} Q(\theta,P)$ exists.

Our construction exploits the well-known duality between $1-\alpha$ confidence regions and tests of the individual null hypotheses

$$H_\theta : Q(\theta, P) = 0 \text{ for } \theta \in \Theta \tag{8}$$

that control the usual probability of a Type 1 error at level α. The hypotheses H_θ are one-sided in the sense that the alternative hypotheses are understood to be

$$K_\theta : Q(\theta, P) > 0 \text{ for } \theta \in \Theta \ .$$

Typically, Θ will be a subset of Euclidean space, in which case there may be an uncountably infinite number of hypotheses in this family. For each null hypothesis H_θ, a test statistic is required such that large values of the test statistic provide evidence against the null hypothesis. We will use the statistic $a_n \hat{Q}_n(\theta)$ for some sequence $a_n \to \infty$ for this purpose.

The construction of critical values for these tests will be based on subsampling. In order 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 $\alpha \in (0,1)$, define

$$\hat{d}_n(\theta, 1-\alpha) = \inf\{x : \frac{1}{N_n} \sum_{1 \le i \le N_n} I\{a_b \hat{Q}_{n,b,i}(\theta) \le x\} \ge 1-\alpha\} \ . \tag{9}$$

We will test each null hypothesis H_θ by comparing the test statistic $a_n \hat{Q}_n(\theta)$ with the critical value $\hat{d}_n(\theta, 1-\alpha)$. The set of θ values corresponding to the accepted hypotheses from this procedure will form $\mathcal{C}_n$, i.e.,

$$\mathcal{C}_n = \{\theta \in \Theta : a_n \hat{Q}_n(\theta) \le \hat{d}_n(\theta, 1-\alpha)\} \ . \tag{10}$$

We now provide conditions under which $\mathcal{C}_n$ defined by (10) satisfies the desired coverage property (2).

Theorem 3.2 *Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P. Denote by $J_n(\cdot, \theta, P)$ the distribution of $a_n(\hat{Q}_n(\theta) - Q(\theta, P))$. Suppose that for every $\theta \in \Theta_0(P)$*

$$\limsup_{n\to\infty} \sup_{x \in \mathbf{R}} \{J_b(x, \theta, P) - J_n(x, \theta, P)\} \le 0 \ . \tag{11}$$

Then, $\mathcal{C}_n$ defined by (10) satisfies (2).

PROOF: Fix $\theta \in \Theta_0(P)$. By Theorem 3.1(i), we have immediately that

$$\liminf_{n\to\infty} P\{a_n\hat{Q}_n(\theta) \leq \hat{d}_n(\theta, 1-\alpha)\} \geq 1-\alpha \ .$$

Thus, by the definition (10), it follows that (2) holds. ■

Remark 3.2 Because $\binom{n}{b}$ may be large, it is often more practical to use the following approximation to (9). Let B_n be a positive sequence of numbers tending to ∞ 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 et al. (1999) that one may approximate (9) by

$$\inf\{x : \frac{1}{B_n} \sum_{1\leq i\leq B_n} I\{a_b\hat{Q}_{n,b,I_i}(\theta) \leq x\} \geq 1-\alpha\}$$

without affecting the conclusions of Theorem 3.2. ■

We now revisit each of the examples described in Section 2 and provide conditions under which the hypothesis (11) of Theorem 3.2 holds. It follows that for each of these examples $\mathcal{C}_n$ defined by (10) satisfies (2).

Example 3.1 (*One-Sided Mean*) Recall the setup of Example 2.1. Let $a_n = n$ and suppose P is such that the variance $\sigma^2(P)$ exists. We will verify that the required condition (11) holds for any such P. Consider $\theta \in \Theta_0(P)$. We may assume without loss of generality that $\mu(P) = 0$. First note that because $\hat{Q}_n(\theta) \geq 0$, we may restrict attention to $x \geq 0$ in (11). There are two cases to consider: $\theta > 0$ and $\theta = 0$. In the former case, we have for any $\epsilon > 0$ that $\bar{X}_n - \theta < \epsilon - \theta$ with probability approaching one. By choosing ϵ sufficiently small, we therefore have that $a_n(\hat{Q}_n(\theta) - Q(\theta, P)) = n(\bar{X}_n - \theta)_+^2 = 0$ with probability approaching one. Hence, (11) holds trivially. In the latter case, $a_n(\hat{Q}_n(\theta) - Q(\theta, P)) = n(\bar{X}_n)_+^2$. Thus, for $x \geq 0$,

$$J_n(x, \theta, P) = P\{n(\bar{X}_n)_+^2 \leq x\} = P\{\sqrt{n}\bar{X}_n \leq \sqrt{x}\} \ .$$

Since $\sqrt{n}\bar{X}_n \xrightarrow{\mathcal{L}} \Phi_{\sigma(P)}$, where $\Phi_{\sigma(P)}$ is the distribution of a mean-zero normal random variable with variance $\sigma^2(P)$, Polya's Theorem implies that

$$\sup_{x\in\mathbf{R}} |P\{\sqrt{n}\bar{X}_n \leq x\} - \Phi_{\sigma(P)}(x)| \to 0 \ .$$

The desired condition (11) now follows from an appeal to the triangle inequality. ■

Example 3.2 (*Two-Sided Mean*) Recall the setup of Example 2.2. Let $a_n = n$ and suppose P is such that the variances $\sigma_X^2(P)$ and $\sigma_Y^2(P)$ exist and $\mu_X(P) \leq \mu_Y(P)$. Denote by $\sigma_{X,Y}(P)$ the covariance of X_i and Y_i under P. We will verify that the required condition (11) holds for any such P. Consider $\theta \in \Theta_0(P)$. First note, as before, that because $\hat{Q}_n(\theta) \geq 0$, we may restrict attention to $x \geq 0$ in (11). There are two cases to consider: either $\mu_X(P) < \mu_Y(P)$ or $\mu_X(P) = \mu_Y(P)$.

If $\mu_X(P) < \mu_Y(P)$, then we must consider the case in which $\theta \in (\mu_X(P), \mu_Y(P))$ and the case in which $\theta \in \{\mu_X(P), \mu_Y(P)\}$ separately. In the former case, it is easy to see that $a_n(\hat{Q}_n(\theta) - Q(\theta, P)) = n(\bar{X}_n - \theta)_+^2 + n(\theta - \bar{Y}_n)_+^2 = 0$ with probability approaching one. Hence, (11) holds trivially. In the latter case, we may assume without loss of generality that $\theta = \mu_X(P)$ and $\mu_X(P) = 0$; the case in which $\theta = \mu_Y(P)$ is symmetric. In this case, $a_n(\hat{Q}_n(\theta) - Q(\theta, P)) = n(\bar{X}_n)_+^2 + n(-\bar{Y}_n)_+^2$. Because $\mu_Y(P) > 0$, we have that with probability tending to 1, $a_n(\hat{Q}_n(\theta) - Q(\theta, P)) = n(\bar{X}_n)_+^2$. The desired condition (11) now follows from the reasoning given in Example 3.1.

If $\mu_X(P) = \mu_Y(P)$, then we may assume without loss of generality that this common value is 0. Then, $a_n(\hat{Q}_n(\theta) - Q(\theta, P)) = n(\bar{X}_n)_+^2 + n(-\bar{Y}_n)_+^2$. Thus, for $x \geq 0$,

$$J_n(x, \theta, P) = P\{n(\bar{X}_n)_+^2 + n(-\bar{Y}_n)_+^2 \leq x\} = P\{(\sqrt{n}\bar{X}_n, -\sqrt{n}\bar{Y}_n)' \in S_x\} ,$$

where S_x is the appropriate convex set. Note that

$$(\sqrt{n}\bar{X}_n, -\sqrt{n}\bar{Y}_n)' \xrightarrow{\mathcal{L}} \Phi_{\sigma_X(P), \sigma_Y(P), -\sigma_{X,Y}(P)} ,$$

where $\Phi_{\sigma_X(P), \sigma_Y(P), -\sigma_{X,Y}(P)}$ is a mean-zero bivariate normal distribution with variances $\sigma_X^2(P)$ and $\sigma_Y^2(P)$ and covariance $-\sigma_{X,Y}(P)$. It follows from Theorem 2.11 of Bhattacharya and Rao (1976) that

$$\sup_{S \in \mathcal{S}} |P\{(\sqrt{n}\bar{X}_n, -\sqrt{n}\bar{Y}_n)' \in S\} - \Phi_{\sigma_X(P), \sigma_Y(P), -\sigma_{X,Y}(P)}(S)| \to 0 ,$$

where $\mathcal{S}$ is the set of all convex subsets of $\mathbf{R}^2$. Hence, the desired condition (11) again follows from an appeal to the triangle inequality. ■

Example 3.3 (*Regression with Interval Outcomes*) Recall the setup of Example 2.3. Let $a_n = n$ and let $\{x_1, \dots, x_J\}$ be a set of vectors in $\mathbf{R}^k$ whose span is of dimension k.

Suppose P is such that (i) $\text{supp}_P(X_i) = \{x_1, \ldots, x_J\}$ and (ii) the variances of Y_1 and Y_2, $\sigma_1^2(P)$ and $\sigma_2^2(P)$, exist. For $l \in \{1, 2\}$ and $j \in \{1, \ldots, J\}$, let $\tau_l(x_j) = E_P\{Y_{li}|X_i = x_j\}$ and

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

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

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

Note that it follows from the above assumptions on P that the vector whose components are given by $\sqrt{n}(\hat{\tau}_l(x_j) - \tau_l(x_j))$ for $j = 1, \ldots, J$ converges in distribution to a multivariate normal distribution.

We will now verify that the required condition (11) holds for this setup. Consider $\theta \in \Theta_0(P)$. First note, as before, that because $\hat{Q}_n(\theta) \geq 0$, we may restrict attention to $x \geq 0$ in (11). There are two cases to consider: either $\theta \in \text{int}(\Theta_0(P))$ or $\theta \in \partial\Theta_0(P)$. In the former case, $\tau_1(x_j) < x_j'\theta < \tau_2(x_j)$ for all j. Thus, it is easy to see that $a_n\hat{Q}_n(\theta) = 0$ with probability tending to 1. Hence, (11) holds trivially. In the latter case, let I_l denote the set of j indices such that $\tau_l(x_j) = x_j'\theta$. Note that at most one of I_1 or I_2 may be empty. Then, with probability approaching one, we have that

$$a_n\hat{Q}_n(\theta) = \sum_{j \in I_1} \frac{n(x_j)}{n} (\sqrt{n}(\hat{\tau}_1(x_j) - x_j'\theta))_+^2 + \sum_{j \in I_2} \frac{n(x_j)}{n} (\sqrt{n}(x_j'\theta - \hat{\tau}_2(x_j)))_+^2 \ .$$

Let $\hat{W}_n(\theta)$ denote the vector whose components are given by $\sqrt{n}(\hat{\tau}_1(x_j) - x_j'\theta)$ for $j \in I_1$ and $\sqrt{n}(x_j'\theta - \hat{\tau}_2(x_j))$ for $j \in I_2$ and let $W(\theta)$ denote the limiting multivariate normal distribution of $\hat{W}_n(\theta)$. For $x \geq 0$, we have that

$$J_n(x, \theta, P) = P\{\hat{W}_n(\theta) \in \hat{S}_x\} \ ,$$

where $\hat{S}_x$ is the appropriate (random) convex set. It follows from Theorem 2.11 of Bhattacharya and Rao (1976) that

$$\sup_{S \in \mathcal{S}} |P\{\hat{W}_n(\theta) \in S\} - P\{W(\theta) \in S\}| \to 0 \ ,$$

where $\mathcal{S}$ is the set of all convex subsets of $\mathbf{R}^{|I_1|+|I_2|}$. Hence, the desired condition (11) again follows from an appeal to the triangle inequality. ■

Example 3.4 (*Moment Inequalities*) Recall the setup of Example 2.4. Let $a_n = n$ and suppose P is such that the variance of the vector whose components are given by $g_j(X_i, \theta)$ for $j = 1, \ldots, m$ exists for all $\theta \in \Theta$. It follows from this assumption that the vector whose components are given by

$$\frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (g_j(X_i, \theta) - E_P\{g_j(X_i, \theta)\})$$

for $j = 1, \ldots, m$ converges in distribution to a multivariate normal random variable for each $\theta \in \Theta$. We will verify that the required condition (11) holds for any such P.

Consider $\theta \in \Theta_0(P)$. There are two cases to consider: either $\theta \in \text{int}(\Theta_0(P))$ or $\theta \in \partial\Theta_0(P)$. In the former case, $E_P\{g_j(X_i, \theta)\} < 0$ for all j, so $a_n\hat{Q}_n(\theta) = 0$ with probability tending to 1. Hence, (11) holds trivially. In the latter case, let I denote the set of j indices such that $E_P\{g_j(X_i, \theta)\} = 0$. Note that I must be non-empty. Then, with probability approaching one, we have that

$$a_n\hat{Q}_n(\theta) = \sum_{j \in I} (\frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} g_j(X_i, \theta))_+^2 \ .$$

Let $\hat{W}_n(\theta)$ denote the vector whose components are given by

$$\frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} g_j(X_i, \theta)$$

for $j \in I$ and let $W(\theta)$ denote the limiting multivariate normal distribution. For $x \geq 0$, we have that

$$J_n(x, \theta, P) = P\{\hat{W}_n(\theta) \in S_x\} \ ,$$

where S_x is the appropriate convex set. It follows from Theorem 2.11 of Bhattacharya and Rao (1976) that

$$\sup_{S \in \mathcal{S}} |P\{\hat{W}_n(\theta) \in S\} - P\{W(\theta) \in S\}| \to 0 \ ,$$

where $\mathcal{S}$ is the set of all convex subsets of $\mathbf{R}^{|I|}$. Hence, the desired condition (11) again follows from an appeal to the triangle inequality. ■

Remark 3.3 In each of the examples from Section 2, there are many other choices of $Q(\theta, P)$ sharing the feature that

$$\Theta_0(P) = \arg\min_{\theta \in \Theta} Q(\theta, P) \ . \tag{12}$$

Consider Example 2.4. Let $g(x, \theta)$ denote the vector whose components are given by $g_j(x, \theta), j = 1, \ldots, m$ and extend the definition of $(\cdot)_+$ on the real line to vectors in the natural way. It is easy to see that

$$Q(\theta, P) = (E_P\{g(X, \theta)\})'_+ W(\theta)(E_P\{g(X, \theta))_+\} ,$$

where $W(\theta)$ is a positive definite matrix for each $\theta \in \Theta$ also satisfies (12) for this example. The objective function given by

$$Q(\theta, P) = \max_{1 \le j \le m} w_j(\theta) E_P\{g_j(X, \theta)\} ,$$

where $w_j(\theta), j = 1, \ldots, m$ are positive for each $\theta \in \Theta$, also satisfies (12) for this example. Each of these choices of $Q(\theta, P)$ will lead to different test statistics $a_n \hat{Q}_n(\theta)$, which in turn will lead to different tests of the individual null hypotheses in (8). Of course, more powerful tests will reject more false hypotheses and thereby lead to smaller confidence regions, so the choice of $Q(\theta, P)$ is an important direction for future research. ■

Remark 3.4 If it were true that under the null hypothesis H_θ, the statistic $a_n \hat{Q}_n(\theta)$ converged in distribution to a random variable that was continuous at its $1-\alpha$ quantile, then it would follow immediately from Theorem 2.2.1 of Politis et al. (1999) that $\mathcal{C}_n$ defined by (10) satisfies (2). Unfortunately, in the examples from Section 2, this assertion does not always hold. Indeed, in many instances, for $\theta \in \text{int}(\Theta_0(P))$, it is true that $a_n \hat{Q}_n(\theta) = 0$ with probability tending to 1. For this reason, it is necessary to appeal to Theorem 3.1 to establish the desired conclusions. ■

Remark 3.5 Chernozhukov et al. (2004) consider the problem of constructing confidence regions $\mathcal{C}_n$ satisfying (2). Their construction relies upon an initial estimate of $\Theta_0(P)$, which they obtain as $\hat{\Theta}_{0,n} = \{\theta \in \Theta : \hat{Q}_n(\theta) < \epsilon_n\}$ where ϵ_n is a positive sequence of constants tending to 0 slowly. Because of this rate restriction, they are able to show that $\hat{\Theta}_{0,n}$ satisfies

$$P\{\Theta_0(P) \subseteq \hat{\Theta}_{0,n}\} \to 1 .$$

They define their confidence region to be

$$\mathcal{C}_n = \{\theta \in \Theta : a_n \hat{Q}_n(\theta) \le \sup_{\theta \in \hat{\Theta}_{0,n}} \hat{d}_n(\theta, 1-\alpha)\} ,$$

where $\hat{d}_n(\theta, 1-\alpha)\}$ is given by (9). Of course, such a construction is clearly more conservative than the one given by (10) whenever both are valid in the following sense. Since $\Theta_0(P) \subseteq \hat{\Theta}_{0,n}$ with probability tending to 1, we have for any $\theta' \in \Theta_0(P)$ that

$$\hat{d}_n(\theta', 1-\alpha) \leq \sup_{\theta \in \Theta_0(P)} \hat{d}_n(\theta, 1-\alpha) \leq \sup_{\theta \in \hat{\Theta}_{0,n}} \hat{d}_n(\theta, 1-\alpha) \tag{13}$$

with probability tending to 1. Often, the first inequality in (13) will be strict for most $\theta \in \Theta_0(P)$. To illustrate this, recall the setup of Example 2.2. Suppose further that $\mu_X(P) < \mu_Y(P)$ and $\sigma_X^2(P) < \sigma_Y^2(P) < \infty$. Then, for $\theta \in \text{int}(\Theta_0(P))$, it follows from Lemma 5.2 in the appendix that $\hat{d}_n(\theta, 1-\alpha) \stackrel{P}{\to} 0$. We also have that $\hat{d}_n(\mu_X(P), 1-\alpha) \stackrel{P}{\to} \sigma_X^2(P) z_{1-\alpha}^2$ and $\hat{d}_n(\mu_Y(P), 1-\alpha) \stackrel{P}{\to} \sigma_Y^2(P) z_{1-\alpha}^2$. It follows that with probability tending to 1, $\hat{d}_n(\mu_X(P), 1-\alpha) < \hat{d}_n(\mu_Y(P), 1-\alpha)$, and so

$$\sup_{\theta \in \Theta_0(P)} \hat{d}_n(\theta, 1-\alpha) = \hat{d}_n(\mu_Y(P), 1-\alpha) \ .$$

Therefore, the first inequality in (13) is strict with probability tending to 1 except when $\theta' = \mu_Y(P)$. In fact, for any $\theta' \neq \mu_Y(P)$, the difference between the two critical values is asymptotically bounded away from zero. ■

Remark 3.6 In certain applications, it may be of interest to test the null hypothesis that there exists some point $\theta \in \Theta$ for which $Q(\theta, P) = 0$. Such a test may be viewed as a specification test of the model. One may test such a hypothesis by first constructing a confidence region for identifiable parameters and then rejecting the hypothesis if the confidence region is empty. To see this, note that under the null hypothesis, $\Theta_0(P) \neq \emptyset$. Thus, there is some $\theta' \in \Theta_0(P)$. Under the assumptions of Theorem 3.2, it must be the case that

$$\liminf_{n \to \infty} P\{\theta' \in \mathcal{C}_n\} \geq 1 - \alpha \ .$$

Thus, under the null hypothesis $\mathcal{C}_n$ will be nonempty with probability at least $1-\alpha$ asymptotically. ■

Remark 3.7 Our construction of critical values for testing $H_\theta : Q(\theta, P) = 0$ versus $K_\theta : Q(\theta, P) > 0$ has used subsampling. We now show by example that the bootstrap may fail to approximate the distribution of the test statistic, $a_n \hat{Q}_n(\theta)$, under the null hypothesis.

To this end, recall the setup of Example 2.1. Let $a_n = n$ and suppose that $P = N(0,1)$. Note that $Q(\theta, P) = 0$ at $\theta = 0$, so the null hypothesis is true when $\theta = 0$. Our argument follows the one given by Andrews (2000) to show that the bootstrap is inconsistent when the parameter is on the boundary of the parameter space. Note that $a_n \hat{Q}_n(0) = (\sqrt{n}\bar{X}_n)^2_+ \sim (Z)^2_+$, where Z is a standard normal random variable. Denote by X_i^*, $i = 1, \ldots, n$ an i.i.d. sequence of random variables with distribution $\hat{P}_n$ given by the empirical distribution of the original observations X_i, $i = 1, \ldots, n$. Define

$$\hat{Q}_n^*(\theta) = (\bar{X}_n^* - \theta)^2_+$$

and consider the event

$$A_c = \{\omega : \liminf_{n\to\infty} \sqrt{n}\bar{X}_n < -c\} \tag{14}$$

for $c \in (0, \infty)$. The Law of the Iterated Logarithm asserts that

$$\limsup_{n\to\infty} \frac{\sqrt{n}\bar{X}_n}{\sqrt{2\log\log n}} = 1 \quad P\text{−a.s.} \ .$$

This in turn implies that

$$\{\sqrt{n}\bar{X}_n > \sqrt{2\log\log n} \ \ \text{i.o.}\} \quad P\text{−a.s.}$$

and by symmetry that

$$\{\sqrt{n}\bar{X}_n < -\sqrt{2\log\log n} \ \ \text{i.o.}\} \quad P\text{−a.s.}$$

It follows that $P\{A_c\} = 1$.

Now consider the bootstrap approximation to the distribution of $a_n\hat{Q}_n(\theta)$ at $\theta = 0$, which is given by

$$\mathcal{L}(a_n(\hat{Q}_n^*(0) - \hat{Q}_n(0))|X_i, i = 1, \ldots, n) \ . \tag{15}$$

This approximation mimics the hypothesis that $Q(0, P) = 0$ by centering $\hat{Q}_n^*(0)$ about $\hat{Q}_n(0)$. For $\omega \in A_c$, consider a subsequence n_k of $n \geq 1$ for which $\sqrt{n_k}\bar{X}_{n_k}(\omega) < -c$ for all k. For such a subsequence, we have, conditionally on $X_i(\omega), i = 1, \ldots, n$, that

$$\begin{aligned}
n_k(\hat{Q}_{n_k}^*(0) - \hat{Q}_{n_k}(0)(\omega)) &= (\sqrt{n_k}\bar{X}_{n_k}^*)^2_+ \\
&= (\sqrt{n_k}(\bar{X}_{n_k}^* - \bar{X}_{n_k}(\omega)) + \sqrt{n_k}\bar{X}_{n_k}(\omega))^2_+ \\
&\leq (\sqrt{n_k}(\bar{X}_{n_k}^* - \bar{X}_{n_k}(\omega)) - c)^2_+ \\
&\stackrel{\mathcal{L}}{\to} (Z - c)^2_+ \ ,
\end{aligned}$$

where Z is again a standard normal random variable. The first equality and the inequality follow from the fact that $\sqrt{n_k}\bar{X}_{n_k}(\omega) < -c < 0$. It follows that the bootstrap fails to approximate the distribution of $a_n\hat{Q}_n(\theta)$ at $\theta = 0$.

It is worthwhile to consider the actual probability of a Type 1 error if the bootstrap approximation above is used instead of subsampling. To this end, first note that the probability of such an error is given by

$$P\{a_n\hat{Q}_n(0) > \hat{c}_{\text{boot},n}(1-\alpha)\} \ , \tag{16}$$

where $\hat{c}_{\text{boot},n}(1-\alpha)$ is the $1-\alpha$ quantile of (15). Recall that $a_n\hat{Q}_n(0) = (\sqrt{n}\bar{X}_n)^2_+ \sim (Z)^2_+$, where Z is a standard normal random variable. Note that

$$a_n(\hat{Q}^*_n(0) - \hat{Q}_n(0)) = (\sqrt{n}(\bar{X}^*_n - \bar{X}_n) + \sqrt{n}\bar{X}_n)^2_+ - (\sqrt{n}\bar{X}_n)^2_+ \ .$$

Therefore, (15) coverges in distribution to

$$\mathcal{L}((Z^* + Z)^2_+ - (Z)^2_+|Z) \ , \tag{17}$$

where Z^* is a standard normal random variable distributed independently of Z. The probability (16) is therefore asymptotically equal to

$$P\{(Z)^2_+ > c(1-\alpha|Z)\} \ ,$$

where $c(1-\alpha|Z)$ is the $1-\alpha$ quantile of (17). Using this representation, we simulate the asymptotic value of (16) and find that the probability of Type 1 error is controlled, but too conservatively for practical purposes. In fact, with 2000 simulations, we estimate that (16) is asymptotically equal to 0 to three significant digits for both $\alpha = .1$ and $\alpha = .05$. ■

3.3 Uniform Consistency in Level

We now turn our attention to providing conditions under which $\mathcal{C}_n$ defined by (10) satisfies (3). Recall from the discussion in the introduction that inferences made on the basis of asymptotic arguments with confidence regions that are not uniformly consistent in level may be very misleading in the sense that asymptotic approximations may be poor for arbitrarily large sample sizes.

Theorem 3.3 *Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P. Denote by $J_n(\cdot, \theta, P)$ the distribution of $a_n(\hat{Q}_n(\theta) - Q(\theta, P))$. Suppose that*

$$\limsup_{n \to \infty} \sup_{\theta \in \Theta} \sup_{P \in \mathbf{P}: \theta \in \Theta_0(P)} \sup_{x \in \mathbf{R}} \{J_b(x, \theta, P) - J_n(x, \theta, P)\} \leq 0 \ . \tag{18}$$

Then, $\mathcal{C}_n$ defined by (10) satisfies (3).

PROOF: By Theorem 3.1(iv), we have immediately that

$$\liminf_{n \to \infty} \inf_{\theta \in \Theta} \inf_{P \in \mathbf{P}: \theta \in \Theta_0(P)} P\{a_n \hat{Q}_n(\theta) \leq \hat{d}_n(\theta, 1 - \alpha)\} \geq 1 - \alpha \ .$$

Thus, by the definition (10), it follows that (3) holds. ∎

We now provide two applications of Theorem 3.3 to constructing confidence regions satisfying the coverage requirement (3).

Example 3.5 (*One-Sided Mean*) Recall the setup of Example 2.1. We will now use Theorem 3.3 to show that for this example $\mathcal{C}_n$ defined by (10) satisfies (3) for a large class of distributions $\mathbf{P}$. To this end, let $a_n = n$ and let $\mathbf{P}$ be a set of distributions satisfying

$$\lim_{\lambda \to \infty} \sup_{P \in \mathbf{P}} E_P \left[\frac{|X - \mu(P)|^2}{\sigma^2(P)} I \left\{ \frac{|X - \mu(P)|}{\sigma(P)} > \lambda \right\} \right] = 0 \ . \tag{19}$$

We will argue by contradiction that the required condition (18) holds. If the result were false, then there would exist a subsequence n_j and a corresponding sequence $(\theta_{n_j}, P_{n_j}) \in \Theta \times \mathbf{P}$ such that $\mu(P_{n_j}) \leq \theta_{n_j}$ and

$$\sup_{x \in \mathbf{R}} \{J_{b_{n_j}}(x, \theta_{n_j}, P_{n_j}) - J_{n_j}(x, \theta_{n_j}, P_{n_j})\} \to \delta \tag{20}$$

for some $\delta > 0$. Since $a_n \hat{Q}_n(\theta_{n_j}) \geq 0$, we may restrict attention to $x \geq 0$. We may thus rewrite (20) as

$$\sup_{x \geq 0} \{P_{n_j}\{(Z_{b_{n_j}}(P_{n_j}) + T_{b_{n_j}}(\theta_{n_j}, P_{n_j}))_+^2 \leq x\} - P_{n_j}\{(Z_{n_j}(P_{n_j}) + T_{n_j}(\theta_{n_j}, P_{n_j}))_+^2 \leq x\}\} \ ,$$

where

$$\begin{aligned} Z_m(P) &= \sqrt{m}(\bar{X}_m - \mu(P)) \\ T_m(\theta, P) &= \sqrt{m}(\mu(P) - \theta) \ . \end{aligned}$$

Since $b_{n_j} \leq n_j$ and $\mu(P_{n_j}) \leq \theta_{n_j}$, we have that $T_{b_{n_j}}(\theta_{n_j}, P_{n_j}) \geq T_{n_j}(\theta_{n_j}, P_{n_j})$. Thus, (20) is bounded above by

$$\sup_{x \geq 0}\{P_{n_j}\{(Z_{b_{n_j}}(P_{n_j}) + T_{n_j}(\theta_{n_j}, P_{n_j}))_+^2 \leq x\} - P_{n_j}\{(Z_{n_j}(P_{n_j}) + T_{n_j}(\theta_{n_j}, P_{n_j}))_+^2 \leq x\}\} \ .$$

By redefining x appropriately and absorbing the term $T_{n_j}(\theta_{n_j}, P_{n_j})$ into x, it follows that

$$\sup_{x \in \mathbf{R}}\{P_{n_j}\{Z_{b_{n_j}}(P_{n_j}) \leq x\} - P_{n_j}\{Z_{n_j}(P_{n_j}) \leq x\}\} \not\to 0 \ .$$

By Lemma 11.4.1 of Lehmann and Romano (2005), this yields the desired contradiction.
■

Example 3.6 (*Two-Sided Mean*) Recall the setup of Example 2.2. As in Example 3.5, we may use Theorem 3.3 to show that for this example $\mathcal{C}_n$ defined by (10) satisfies (3) for a large class of distributions $\mathbf{P}$. To this end, let $a_n = n$, let $\mathbf{P}$ be a set of bivariate distributions such that the marginal distributions satisfy (19) and $\mu_X(P) \leq \mu_Y(P)$. As before, we will argue by contradiction that the required condition (18) holds, but we will require the following result, which generalizes Lemma 11.4.1 of Lehmann and Romano (2005), in order to do so. Again, for the sake of continuity, the proof of the result is found in the appendix.

Lemma 3.1 *Let $\mathbf{P}$ be a set of bivariate distributions such that the marginal distributions satisfy (19). Let $(X_i, Y_i), i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ and denote by $\Phi_{\sigma_X(P), \sigma_Y(P), \rho(P)}(\cdot)$ the probability distribution of a bivariate normal random variable with mean 0, standard deviations $\sigma_X(P)$ and $\sigma_Y(P)$, and correlation $\rho(P)$. Then,*

$$\sup_{P \in \mathbf{P}} \sup_{S \in \mathcal{S}} |P\{(\sqrt{n}(\bar{X}_n - \mu_X(P)), \sqrt{n}(\bar{Y}_n - \mu_Y(P)))' \in S\} - \Phi_{\sigma_X(P), \sigma_Y(P), \rho(P)}(S)\}$$

converges to 0, where

$$\mathcal{S} = \{S \subseteq \mathbf{R}^2 : S \text{ convex and } \Phi_{1,1,1}(\partial S) = \Phi_{1,1,-1}(\partial S) = 0\} \ . \tag{21}$$

PROOF: See appendix. ■

We now return to verifying (18). If the result were false, then there would exist a subsequence n_j and a corresponding sequence $(\theta_{n_j}, P_{n_j}) \in \Theta \times \mathbf{P}$ such that $\mu_X(P_{n_j}) \leq \theta_{n_j} \leq \mu_Y(P_{n_j})$ and

$$\sup_{x \in \mathbf{R}} \{J_{b_{n_j}}(x, \theta_{n_j}, P_{n_j}) - J_{n_j}(x, \theta_{n_j}, P_{n_j})\} \to \delta \tag{22}$$

for some $\delta > 0$. Note that $a_{b_{n_j}} \hat{Q}_{b_{n_j}}(\theta_{n_j})$

$$(Z_{1,b_{n_j}}(P_{n_j}) + T_{1,b_{n_j}}(\theta_{n_j}, P_{n_j}))_+^2 + (Z_{2,b_{n_j}}(P_{n_j}) + T_{2,b_{n_j}}(\theta_{n_j}, P_{n_j}))_+^2 ,$$

where

$$\begin{aligned} Z_{1,m}(P) &= \sqrt{m}(\bar{X}_m - \mu_X(P)) \\ Z_{2,m}(P) &= \sqrt{m}(\mu_Y(P) - \bar{Y}_m) \\ T_{1,m}(\theta, P) &= \sqrt{m}(\mu_X(P) - \theta) \\ T_{2,m}(\theta, P) &= \sqrt{m}(\theta - \mu_Y(P)) . \end{aligned}$$

Let $\tilde{J}_{b_{n_j}}(x, \theta_{n_j}, P_{n_j})$ be the distribution of the statistic

$$(Z_{1,b_{n_j}}(P_{n_j}) + T_{1,n_j}(\theta_{n_j}, P_{n_j}))_+^2 + (Z_{2,b_{n_j}}(P_{n_j}) + T_{2,n_j}(\theta_{n_j}, P_{n_j}))_+^2 .$$

Since $T_{k,b_{n_j}}(\theta_{n_j}, P_{n_j}) \geq T_{k,n_j}(\theta_{n_j}, P_{n_j})$, we have that

$$\tilde{J}_{b_{n_j}}(x, \theta_{n_j}, P_{n_j}) \geq J_{b_{n_j}}(x, \theta_{n_j}, P_{n_j}) .$$

Thus, (22) implies that

$$\sup_{x \in \mathbf{R}} \{\tilde{J}_{b_{n_j}}(x, \theta_{n_j}, P_{n_j}) - J_{n_j}(x, \theta_{n_j}, P_{n_j})\} \not\to 0 . \tag{23}$$

Since $a_n \hat{Q}_n(\theta) \geq 0$, we may restrict attention to $x \geq 0$. For such x, $\tilde{J}_{b_{n_j}}(x, \theta_{n_j}, P_{n_j})$ is simply the probability under P_{n_j} that the vector $(Z_{1,b_{n_j}}(P_{n_j}), Z_{2,b_{n_j}}(P_{n_j}))'$ lies in a set $S_x \in \mathcal{S}$, where $\mathcal{S}$ is defined by (21). Importantly, $J_{n_j}(x, \theta_{n_j}, P_{n_j})$ is simply the probability under P_{n_j} that the vector $(Z_{1,n_j}(P_{n_j}), Z_{2,n_j}(P_{n_j}))'$ lies in the same set. But, by Lemma 3.1, however, we know that

$$\sup_{S \in \mathcal{S}} |P_{n_j}\{(Z_{1,n_j}(P_{n_j}), Z_{2,n_j}(P_{n_j}))' \in S\} - \Phi_{\sigma_X(P_{n_j}), \sigma_Y(P_{n_j}), \rho(P_{n_j})}(S)| \to 0 ,$$

and

$$\sup_{S \in \mathcal{S}} |P_{n_j}\{(Z_{1,b_{n_j}}(P_{n_j}), Z_{2,b_{n_j}}(P_{n_j}))' \in S\} - \Phi_{\sigma_X(P_{n_j}), \sigma_Y(P_{n_j}), \rho(P_{n_j})}(S)| \to 0 .$$

An appeal to the triangle inequality yields the desired contradiction to (23). ■

Remark 3.8 Recall that Imbens and Manski (2004) analyze a special case of Example 2.2 in which $X_i = W_i D_i$ and $Y_i = W_i D_i + 1 - D_i$ where $W_i \in [0,1]$ and $D_i \in \{0,1\}$. The motivation for their study of this problem stems from considering

$$\mathcal{C}_n = [\bar{X}_n - \frac{z_{1-\alpha}\hat{\sigma}_{X,n}}{\sqrt{n}}, \bar{Y}_n + \frac{z_{1-\alpha}\hat{\sigma}_{Y,n}}{\sqrt{n}}] \ , \tag{24}$$

where $\hat{\sigma}_{X,n}$ and $\hat{\sigma}_{Y,n}$ are the usual consistent estimators of $\sigma_X(P)$ and $\sigma_Y(P)$ respectively, and $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. For any P such that $P\{D_i = 1\} < 1$, $\mathcal{C}_n$ defined by (24) satisfies (2). To see this, consider the two cases $\theta \in \text{int}(\Theta_0(P))$ and $\theta \in \partial\Theta_0(P)$ separately. In the former case, $\bar{X}_n < \theta < \bar{Y}_n$ with probability approaching 1, which implies that $\theta \in \mathcal{C}_n$ with probability approaching 1 as well. In the latter case, $\theta = \mu_X(P)$ or $\theta = \mu_Y(P)$. Suppose $\theta = \mu_X(P)$; the case in which $\theta = \mu_Y(P)$ is completely symmetric. Then, $\theta < \bar{Y}_n$ with probability approaching 1 and $\theta > \bar{X}_n - \frac{z_{1-\alpha}\hat{\sigma}_{X,n}}{\sqrt{n}}$ with probability at least $1-\alpha$ asymptotically. We therefore have that $\theta \in \mathcal{C}_n$ with probability at least $1-\alpha$ asymptotically, as desired. Imbens and Manski (2004) show, however, that this convergence is not uniform over all P satisfying $P\{D_i = 1\} < 1$. Specifically, they show that for every sample size n there is a P with $P\{D_i = 1\}$ sufficiently close, but not equal, to 1 and a $\theta \in \Theta_0(P)$ for which

$$P\{\theta \in \mathcal{C}_n\} \approx 1 - 2\alpha \ .$$

To rectify this shortcoming, Imbens and Manski (2004) propose an alternative to $\mathcal{C}_n$ defined by (24) which satisfies the uniform coverage requirement (3) for $\mathbf{P} = \mathbf{P}'$ such that

$$\begin{aligned} \inf_{P \in \mathbf{P}'} P\{D_i = 1\} &> 0 \\ \inf_{P \in \mathbf{P}'} \sigma^2_{W_i|D_i=1}(P) &> 0 \ . \end{aligned}$$

We may apply the analysis of Example 3.6 and conclude immediately that $\mathcal{C}_n$ defined by (10) also satisfies (3) for $\mathbf{P}$ such that the distributions of X_i and Y_i defined above satisfy (19). This class of distributions is larger than the one considered by Imbens and Manski (2004). To see this, first note that

$$\begin{aligned} \inf_{P \in \mathbf{P}'} \sigma^2_{W_i D_i}(P) &\geq \inf_{P \in \mathbf{P}'} P\{D_i = 1\}\sigma^2_{W_i|D_i=1}(P) \\ &\geq \inf_{P \in \mathbf{P}'} P\{D_i = 1\} \inf_{P \in \mathbf{P}'} \sigma^2_{W_i|D_i=1}(P) > 0 \ . \end{aligned}$$

Since W_iD_i is supported on a compact set, it follows that for any $\delta > 0$

$$\sup_{P \in \mathbf{P}'} E_P \left\{ \left(\frac{|W_iD_i - \mu_{W_iD_i}(P)|}{\sigma_{W_iD_i}(P)} \right)^{2+\delta} \right\} < \infty \ .$$

From the pointwise inequality

$$x^2 I\{x > \lambda\} \leq \frac{|x|^{2+\delta}}{\lambda^\delta}$$

for $\lambda > 0$ and $\delta > 0$, we therefore have that (19) holds for W_iD_i and $\mathbf{P} = \mathbf{P}'$. Next note that to show that

$$\inf_{P \in \mathbf{P}'} \sigma^2_{W_iD_i + 1 - D_i}(P) > 0$$

it suffices to show that

$$\inf_{P \in \mathbf{P}'} \rho_{W_iD_i, 1-D_i}(P) > -1 \ . \tag{25}$$

But, by direct calculation, we have that

$$\rho_{W_iD_i, 1-D_i}(P) = \frac{-1}{\sqrt{1 + \frac{\sigma^2_{W_i|D_i=1}(P)}{E_P\{W_i|D_i=1\}^2 P\{D_i=0\}}}} \ .$$

It now follows immediately that (25) holds. We may therefore argue as above for W_iD_i to establish that (19) holds for $W_iD_i + 1 - D_i$ and $\mathbf{P} = \mathbf{P}'$. Hence, $\mathbf{P}'$ is a subset of the set of distributions obtained by applying the analysis of Example 3.6.

It is worthwhile to note that the shortcoming of the construction given by (24) pointed out by Imbens and Manski (2004) disappears if we require further that the confidence region $\mathcal{C}_n = [\hat{L}_n, \hat{U}_n]$ is *equitailed* in the sense that for any $\theta \in \Theta_0(P)$

$$\begin{aligned} \limsup_{n \to \infty} P\{\theta < \hat{L}_n\} &\leq \frac{\alpha}{2} \\ \limsup_{n \to \infty} P\{\theta > \hat{U}_n\} &\leq \frac{\alpha}{2} \ . \end{aligned}$$

One such confidence region is given by (24) in which $z_{1-\alpha}$ is replaced by $z_{1-\frac{\alpha}{2}}$. Note that even if $P\{D_i = 1\} = 1$, so $\Theta_0(P)$ is a singleton, the confidence region $\mathcal{C}_n$ defined in this way satisfies (2). ■

3.4 Confidence Regions for Functions of Identifiable Parameters

In this section, we consider the problem of constructing sets satisfying (5) and (6). Let $f : \Theta \to \Lambda$ be given. Our construction again relies upon test inversion, but in this case the individual null hypotheses are given by

$$H_\lambda : \lambda \in \Lambda_0(P) \text{ for } \lambda \in \Lambda \ , \tag{26}$$

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

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

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

$$\begin{aligned} \lambda \in \Lambda_0(P) &\iff \exists\, \theta \in \Theta_0(P) \text{ s.t. } f(\theta) = \lambda \\ &\iff \exists\, \theta \in \Theta \text{ s.t. } Q(\theta, P) = 0 \text{ and } f(\theta) = \lambda \\ &\iff \exists\, \theta \in f^{-1}(\lambda) \text{ s.t. } Q(\theta, P) = 0 \\ &\iff \inf_{\theta \in f^{-1}(\lambda)} Q(\theta, P) = 0 \ . \end{aligned}$$

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

$$\inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \ , \tag{27}$$

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 the test statistic (27) in place of our earlier test statistic $a_n \hat{Q}_n(\theta)$. For $\alpha \in (0, 1)$, define

$$\hat{d}_n^f(\lambda, 1-\alpha) = \inf\{x : \frac{1}{N_n} \sum_{1 \le i \le N_n} I\{\inf_{\theta \in f^{-1}(\lambda)} a_b \hat{Q}_{n,b,i}(\theta) \le x\} \ge 1 - \alpha\} \ ,$$

and let

$$\mathcal{C}_n^f = \{\lambda \in \Lambda : \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \le \hat{d}_n^f(\lambda, 1-\alpha)\} \ . \tag{28}$$

We now have the following theorem, which generalizes Theorems 3.2 and 3.3.

Theorem 3.4 *Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P. Denote by $J_n(\cdot, \lambda, P)$ the distribution of*

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

(i) Suppose that for every $\lambda \in \Lambda_0(P)$

$$\limsup_{n \to \infty} \sup_{x \in \mathbf{R}} \{J_b(x, \lambda, P) - J_n(x, \lambda, P)\} \leq 0 \ . \tag{29}$$

Then, $\mathcal{C}_n^f$ defined by (28) satisfies (5).

(ii) Suppose that

$$\limsup_{n \to \infty} \sup_{\lambda \in \Lambda} \sup_{P \in \mathbf{P} : \lambda \in \Lambda_0(P)} \sup_{x \in \mathbf{R}} \{J_b(x, \lambda, P) - J_n(x, \lambda, P)\} \leq 0 \ . \tag{30}$$

Then, $\mathcal{C}_n^f$ defined by (28) satisfies (6).

PROOF: The proof is identical to the proofs of Theorems 3.2 and 3.3. ■

We now provide a simple illustration of the use of Theorem 3.4 to construct sets satisfying (5) and (6).

Example 3.7 Consider the following straightforward generalization of Example 2.1. 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 interest focuses on $\theta_{0,1}$ rather than the entire vector θ_0; that is, the object of interest is $f(\theta_0)$, where $f : \mathbf{R}^2 \to \mathbf{R}$ is defined by $f(\theta) = \theta_1$, instead of θ_0. Note that in this instance $\Lambda_0(P)$ is simply $\{\theta_1 \in \mathbf{R} : \theta_1 \geq \mu_X(P)\}$.

First, consider the problem of constructing sets satisfying (5). To this end, let $a_n = n$ and suppose P is such that $\sigma_X^2(P)$ exists. Consider $\theta_1 \in \Lambda_0(P)$. Assume without loss of generality that $\mu_X(P) = 0$. Then,

$$\begin{aligned}\inf_{f^{-1}(\theta_1)} a_n \hat{Q}_n(\theta) &= \inf_{\theta_2 \in \mathbf{R}} n(\bar{X}_n - \theta_1)_+^2 + n(\bar{Y}_n - \theta_2)_+^2 \\ &= n(\bar{X}_n - \theta_1)_+^2 \ .\end{aligned}$$

It now follows immediately from the analysis of Example 3.1 that (29) holds. Therefore, $\mathcal{C}_n^f$ defined by (28) satisfies (5).

Now consider the problem of constructing sets satisfying (6). As before, let $a_n = n$ and let $\mathbf{P}$ be a set of distributions for which the marginal distribution of X satisfies (19). Consider $\theta_1 \in \Lambda_0(P)$. Since $\inf_{f^{-1}(\theta_1)} a_n \hat{Q}_n(\theta)$ is simply $n(\bar{X}_n - \theta_1)_+^2$, it follows immediately from the analysis of Example (3.5) that (30) holds. Therefore, $\mathcal{C}_n^f$ defined by (28) satisfies (6). ■

Remark 3.9 Of course, given a confidence region for identifiable parameters $\mathcal{C}_n$, one crude construction of a confidence region for a function of identifiable parameters is available as the image of $\mathcal{C}_n$ under the function of interest. Unfortunately, such a construction will typically be very conservative. ■

4 Conclusion

This paper has provided computationally intensive, yet feasible methods for inference for a large class of partially identified models. The class of models we have considered are defined by a population objective function $Q(\theta, P)$. The main problem we have considered is the construction of random sets that contain each identifiable parameter with at least some prespecified probability asymptotically. We have also extended these constructions to situations in which the object of interest is the image of an identifiable parameter under a known function. We have verified that these constructions can be applied in several examples of interest.

The results developed in this paper build upon earlier work by Chernozhukov et al. (2004), who also consider the problem of inference for the same class of partially identified models. Our construction of confidence regions for identifiable parameters produces smaller confidence regions than theirs without violating the desired coverage property. This feature persists even asymptotically.

We have also provided conditions under which our confidence regions are uniformly consistent in level. Imbens and Manski (2004) consider the problem of constructing confidence regions for identifiable parameters that are uniformly consistent in level, but their analysis considers only the special case of our 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. Our results therefore provide a generalization of their results to this broader class of models. In order to prove these results, we have first derived more general conditions under which confidence regions for a parameter constructed using subsampling are uniformly consistent in level. These results on subsampling are of interest independently.

In a companion paper, Romano and Shaikh (2006), we consider the problem of constructing random sets that contain the entire identified set with at least some pre-specified probability asymptotically. There we also provide an empirical illustration of the methodologies in both papers.

There are several important directions for future research. Here, we conclude by briefly mentioning two. First, while we have shown that the confidence regions that we have constructed are uniformly consistent in level under weak assumptions, ideally we would like confidence regions with finite-sample validity. Of course, it may not always be possible to construct meaningful confidence regions with finite-sample validity, but it may be possible to do so in specific instances of the general class of models. We hope to explore this possibility in future research. Second, we have not considered the question of whether the confidence regions we have constructed are optimal. Recall from the brief discussion in Remark 3.3 that this question is closely related to the optimal choice of the population objective function $Q(\theta, P)$. Of course, questions of optimality inevitably rely upon a choice of optimality criterion. When the object of interest is an identifiable parameter, this question essentially reduces to the optimal testing of the individual null hypotheses in (8), so several optimality criteria from the theory of testing single hypotheses may be used for this purpose.

5 Appendix: Auxillary Results and Proofs

5.1 Proof of Theorem 3.1.1

We begin with a basic result establishing the relationship between the quantiles of two distribution functions F and G when they are "close" in a certain sense. Recall that for any distribution F on the real line, $F^{-1}(1-\alpha)$ is defined to be $\inf\{x : F(x) \geq 1-\alpha\}$.

Lemma 5.1 *Let F and G be (nonrandom) distribution functions on $\mathbf{R}$. Then, provided that all arguments are strictly between 0 and 1, the following are true:*

(i) If $\sup_{x\in\mathbf{R}}\{G(x)-F(x)\} \leq \epsilon$, then

$$G^{-1}(1-\alpha) \geq F^{-1}(1-(\alpha+\epsilon)) \ .$$

(ii) If $\sup_{x\in\mathbf{R}}\{F(x)-G(x)\} \leq \epsilon$, then

$$G^{-1}(\alpha) \leq F^{-1}(\alpha+\epsilon) \ .$$

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

(iii) If $\sup_{x\in\mathbf{R}}\{G(x)-F(x)\} \leq \epsilon$, then

$$P\{X \leq G^{-1}(1-\alpha)\} \geq 1-(\alpha+\epsilon) \ .$$

(iv) If $\sup_{x\in\mathbf{R}}\{F(x)-G(x)\} \leq \epsilon$, then

$$P\{X \geq G^{-1}(\alpha)\} \geq 1-(\alpha+\epsilon) \ .$$

(v) If $\sup_{x\in\mathbf{R}}|G(x)-F(x)| \leq \epsilon$, then

$$P\{G^{-1}(\alpha) \leq X \leq G^{-1}(1-\alpha)\} \geq 1-2(\alpha+\epsilon) \ .$$

If $\hat{G}$ is a random distribution function on $\mathbf{R}$, then, provided all arguments are between 0 and 1, we have further that:

(vi) *If* $P\{\sup_{x\in\mathbf{R}}\{\hat{G}(x) - F(x)\} \le \epsilon\} \ge 1-\delta$*, then*

$$P\{X \le \hat{G}^{-1}(1-\alpha)\} \ge 1-(\alpha+\epsilon+\delta)\ .$$

(vii) *If* $P\{\sup_{x\in\mathbf{R}}\{F(x) - \hat{G}(x)\} \le \epsilon\} \ge 1-\delta$*, then*

$$P\{X \ge \hat{G}^{-1}(\alpha)\} \ge 1-(\alpha+\epsilon+\delta)\ .$$

(viii) *If* $P\{\sup_{x\in\mathbf{R}}|\hat{G}(x) - F(x)| \le \epsilon\} \ge 1-\delta$*, then*

$$P\{\hat{G}^{-1}(\alpha) \le X \le \hat{G}^{-1}(1-\alpha)\} \ge 1-2(\alpha+\epsilon+\delta)\ .$$

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

To prove (iii), note that because $\sup_{x\in\mathbf{R}}\{G(x)-F(x)\} \le \epsilon$, it follows from (i) that $\{X \le G^{-1}(1-\alpha)\} \subseteq \{X \le F^{-1}(1-(\alpha+\epsilon))\}$. Hence, $P\{X \le G^{-1}(1-\alpha)\} \le P\{X \le F^{-1}(1-(\alpha+\epsilon))\} \ge 1-(\alpha+\epsilon)$. Using the same reasoning, (iv) follows from (ii) and the assumption that $\sup_{x\in\mathbf{R}}\{F(x)-G(x)\} \le \epsilon$. To see (v), note that

$$\begin{aligned} & P\{G^{-1}(\alpha) \le X \le G^{-1}(1-\alpha)\} \\ \ge\ & 1-P\{X < G^{-1}(\alpha)\} - P\{X > G^{-1}(1-\alpha)\} \\ =\ & P\{X \ge G^{-1}(\alpha)\} + P\{X \le G^{-1}(1-\alpha)\} - 1 \\ \ge\ & 1-2(\alpha+\epsilon)\ , \end{aligned}$$

where the first inequality follows from the Bonferroni inequality and the second inequality follows from (iii) and (iv).

To prove (vi), note that

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

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

We now show that the subsampling estimate of the distribution of a statistic based on a sample of size n is uniformly close to the distribution of the statistic based on a sample of size b.

Lemma 5.2 *Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P. Denote by $J_n(\cdot, P)$ the distribution of $\tau_n(\hat{\vartheta}_n - \vartheta(P))$. Suppose $\hat{\vartheta}_n \in \mathbf{R}$. Let $N_n = \binom{n}{b}$, $k_n = \lfloor \frac{n}{b_n} \rfloor$ and define $L_n(x)$ according to (7). Then, for any $\epsilon > 0$, we have that*

$$P\{\sup_{x\in\mathbf{R}} |L_n(x) - J_b(x, P)| > \epsilon\} \leq \frac{1}{\epsilon}\sqrt{\frac{2\pi}{k_n}} . \tag{31}$$

We also have that for any $0 < \delta < 1$

$$P\{\sup_{x\in\mathbf{R}} |L_n(x) - J_b(x, P)| > \epsilon\} \leq \frac{\delta}{\epsilon} + \frac{2}{\epsilon}\exp\{-2k_n\delta^2\} . \tag{32}$$

Proof: Let $\epsilon > 0$ be given and define $S_n(x, P; X_1, \ldots, X_n)$ by

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

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

$$\frac{1}{N_n} \sum_{1 \le i \le N_n} I\{\tau_b(\hat{\vartheta}_{n,b,i} - \vartheta(P)) \le x\} - J_b(x, P)$$

as

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

Note further that

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

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

$$P\{\frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \sup_{x \in \mathbf{R}} |S_n(x, P; X_{\pi(1)}, \ldots, X_{\pi(n)})| > \epsilon\} \ . \tag{33}$$

Using Markov's inequality, (33) can be bounded by

$$\frac{1}{\epsilon} E\{\sup_{x \in \mathbf{R}} |S_n(x, P; X_1, \ldots, X_n)|\} \tag{34}$$

$$= \frac{1}{\epsilon} \int_0^1 P\{\sup_{x \in \mathbf{R}} |S_n(x, P; X_1, \ldots, X_n)| > u\} du \ . \tag{35}$$

We may then use the Dvoretsky-Kiefer-Wolfowitz inequality to bound (35) by

$$\frac{1}{\epsilon} \int_0^1 2 \exp\{-2k_n u^2\} du \ ,$$

which, when evaluated, yields

$$\frac{2}{\epsilon} \sqrt{\frac{2\pi}{k_n}} [\Phi(2\sqrt{k_n}) - \frac{1}{2}] < \frac{1}{\epsilon} \sqrt{\frac{2\pi}{k_n}} \ .$$

To establish (32), note that for any $0 < \delta < 1$, we have that

$$E\{\sup_{x \in \mathbf{R}} |S_n(x, P; X_1, \ldots, X_n)|\} \le \delta + P\{\sup_{x \in \mathbf{R}} |S_n(x, P; X_1, \ldots, X_n)| > \delta\} \ . \tag{36}$$

The result (32) now follows immediately by using the Dvoretsky-Kiefer-Wolfowitz inequality to bound the second term on the right hand side in (36). ■

Using Lemma 5.2, we have the following bounds on the probabilty that the subsampling estimate of the distribution of a statistic based on sample of size n is uniformly close to the distribution of the statistic based on a sample of size n.

Lemma 5.3 *Let $X_i, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P. Denote by $J_n(\cdot, P)$ the distribution of the statistic $\tau_n(\hat{\vartheta}_n - \vartheta(P))$. Suppose $\hat{\vartheta}_n \in \mathbf{R}$. Define $L_n(x)$ according to (7). Then, for all $\epsilon > 0$ and $\gamma \in (0,1)$, we have the following:*

(i) $P\{\sup_{x \in \mathbf{R}}\{L_n(x) - J_n(x, P)\} > \epsilon\} \le$

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

(ii) $P\{\sup_{x \in \mathbf{R}}\{J_n(x, P) - L_n(x)\} > \epsilon\} \le$

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

(iii) $P\{|\sup_{x \in \mathbf{R}}\{L_n(x) - J_n(x, P)\}| > \epsilon\} \le$

$$\frac{1}{\gamma\epsilon}\sqrt{\frac{2\pi}{k_n}} + I\{\sup_{x \in \mathbf{R}}|J_b(x, P) - J_n(x, P)| > (1-\gamma)\epsilon\} \ .$$

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

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

where the final inequality follows from Lemma 5.2. A similar argument establishes (ii) and (iii). ■

We may assemble the above results together to make the following statements about the finite-sample coverage properties of confidence intervals formed using quantiles of the subsampling estimate of the distribution of the statistic based on sample of size n.

Lemma 5.4 *Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$. Denote by $J_n(\cdot, P)$ the distribution of the statistic $\tau_n(\hat{\vartheta}_n - \vartheta(P))$. Suppose $\hat{\vartheta}_n \in \mathbf{R}$. Define $L_n(x)$ according to (7). Let*

$$\begin{aligned}
\delta_{1,n}(\epsilon, \gamma, P) &= \frac{1}{\gamma\epsilon}\sqrt{\frac{2\pi}{k_n}} + I\{\sup_{x \in \mathbf{R}}\{J_b(x, P) - J_n(x, P)\} > (1-\gamma)\epsilon\} \\
\delta_{2,n}(\epsilon, \gamma, P) &= \frac{1}{\gamma\epsilon}\sqrt{\frac{2\pi}{k_n}} + I\{\sup_{x \in \mathbf{R}}\{J_n(x, P) - J_b(x, P)\} > (1-\gamma)\epsilon\} \\
\delta_{3,n}(\epsilon, \gamma, P) &= \frac{1}{\gamma\epsilon}\sqrt{\frac{2\pi}{k_n}} + I\{\sup_{x \in \mathbf{R}}|J_b(x, P) - J_n(x, P)| > (1-\gamma)\epsilon\} \ .
\end{aligned}$$

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

(i) $P\{\tau_n(\hat{\vartheta}_n - \vartheta(P)) \leq L_n^{-1}(1-\alpha)\} \geq$

$$1 - (\alpha + \epsilon + \delta_{1,n}(\epsilon, \gamma, P)) \ .$$

(ii) $P\{\tau_n(\hat{\vartheta}_n - \vartheta(P)) \geq L_n^{-1}(\alpha)\} \geq$

$$1 - (\alpha + \epsilon + \delta_{2,n}(\epsilon, \gamma, P)) \ .$$

(iii) $P\{L_n^{-1}(\frac{\alpha}{2}) \leq \tau_n(\hat{\vartheta}_n - \vartheta(P)) \leq L_n^{-1}(1 - \frac{\alpha}{2})\} \geq$

$$1 - (\alpha + \epsilon + \delta_{3,n}(\epsilon, \gamma, P)) \ .$$

Moreover,

(iv) $\inf_{P \in \mathbf{P}} P\{\tau_n(\hat{\vartheta}_n - \vartheta(P)) \leq L_n^{-1}(1-\alpha)\} \geq$

$$1 - (\alpha + \epsilon + \sup_{P \in \mathbf{P}} \delta_{1,n}(\epsilon, \gamma, P)) \ .$$

(v) $\inf_{P \in \mathbf{P}} P\{\tau_n(\hat{\vartheta}_n - \vartheta(P)) \geq L_n^{-1}(\alpha)\} \geq$

$$1 - (\alpha + \epsilon + \sup_{P \in \mathbf{P}} \delta_{2,n}(\epsilon, \gamma, P)) \ .$$

(vi) $\inf_{P \in \mathbf{P}} P\{L_n^{-1}(\frac{\alpha}{2}) \leq \tau_n(\hat{\vartheta}_n - \vartheta(P)) \leq L_n^{-1}(1 - \frac{\alpha}{2})\} \geq$

$$1 - (\alpha + \epsilon + \sup_{P \in \mathbf{P}} \delta_{3,n}(\epsilon, \gamma, P)) \ .$$

PROOF: Let $\epsilon > 0$ and $\gamma \in (0,1)$ be given. By part (i) of Lemma 5.3, we have that

$$P\{\sup_{x \in \mathbf{R}}\{L_n(x) - J_n(x,P)\} > \epsilon\} \leq \delta_{1,n}(\epsilon, \gamma, P) \ .$$

Assertion (i) follows by applying part (vi) of Lemma 5.1. Assertions (ii) and (iii) are established similarly. Assertions (iv), (v), and (vi) follow immediately by taking the infimum over $P \in \mathbf{P}$ of both sides of inequalities (i), (ii) and (iii). ■

We are now in a position to prove Theorem 3.1, which is an asymptotic version of Lemma 5.4.

PROOF OF THEOREM 3.1: To prove (i), note that by part (i) of Lemma 5.4, we have for any $\epsilon > 0$ and $\gamma \in (0,1)$ that

$$P\{\tau_n(\hat{\theta}_n - \theta(P)) \leq L_n^{-1}(1-\alpha)\} \geq 1 - (\alpha + \epsilon + \delta_{1,n}(\epsilon, \gamma, P)) \ ,$$

where

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

By the assumption on $\sup_{x\in\mathbf{R}}\{J_b(x,P) - J_n(x,P)\}$, we have that for every $\epsilon > 0$, $\delta_{1,n}(\epsilon,\gamma,P) \to 0$. Thus, there exists a sequence $\epsilon_n > 0$ tending to 0 so that $\delta_{1,n}(\epsilon_n, \gamma, P) \to 0$. The desired claim now follows from applying part (i) of Lemma 5.4 to this sequence. The proofs (ii) and (iii) follow in exactly the same way. The proofs of (iv), (v) and (vi) also follow from the same reasoning with the assumption on $\sup_{P\in\mathbf{P}}\sup_{x\in\mathbf{R}}\{J_b(x,P) - J_n(x,P)\}$. ■

5.2 Proof of Lemma 3.3.1

First note that we may assume without loss of generality that $\sigma_X(P) = \sigma_Y(P) = 1$. We will now argue by contradiction. If the result were false, then there would exist a subsequence n_j and a corresponding sequence $P_{n_j} \in \mathbf{P}$ such that $\sigma_X(P_{n_j}) = \sigma_Y(P_{n_j}) = 1$, $\rho(P_{n,j}) \to \rho \in [-1,1]$ and

$$\sup_{S\in\mathcal{S}} |P_{n_j}\{(\sqrt{n_j}(\bar{X}_{n_j} - \mu_X(P_{n_j})), \sqrt{n_j}(\bar{Y}_{n_j} - \mu_Y(P_{n_j})))' \in S\} - \Phi_{1,1,\rho(P_{n_j})}(S)\}|$$

does not converge to zero. There are two cases to consider: either $|\rho| < 1$ or $|\rho| = 1$.

First consider the case in which $|\rho| < 1$. In this case, there exists $\delta > 0$ such that $|\rho| \leq 1 - \delta$. We will use the Cramer-Wold device to establish that

$$(\sqrt{n}(\bar{X}_n - \mu_X(P_{n_j})), \sqrt{n}(\bar{Y}_n - \mu_Y(P_{n_j})))' \xrightarrow{\mathcal{L}} \Phi_{1,1,\rho}$$

under P_{n_j}. Let $(0,0) \neq (a,b) \in \mathbf{R}^2$ and consider

$$\sqrt{n_j}(a\bar{X}_{n_j} + b\bar{Y}_{n_j} - a\mu_X(P_{n_j}) - b\mu_Y(P_{n_j})) \ . \tag{37}$$

To show that (37) converges in distribution to the appropriate normal distribution, by Lemma 11.4.1 of Lehmann and Romano (2005) it suffices to show that $aX + bY$ satisfies (19) when $(X,Y) \sim P \in \mathbf{P}$ and $\sigma_X(P) = \sigma_Y(P) = 1$. We may assume without loss of generality that both a and b are nonzero, since otherwise the result follows from Lemma 11.4.1 of Lehmann and Romano (2005). Moreover, since (aX, bY) has a distribution in $\mathbf{P}$ whenever (X,Y) does, we may assume further that $a = b = 1$. For any $\lambda > 0$ and any such P, note that

$$\begin{aligned} & E_P\left[\frac{(X+Y-\mu_X(P)-\mu_Y(P))^2}{\sigma^2_{X+Y}(P)} I\{\frac{|X+Y-\mu_X(P)-\mu_Y(P)|}{\sigma_{X+Y}(P)} > \lambda\}\right] \\ \leq \ & E_P\left[\frac{(X+Y-\mu_X(P)-\mu_Y(P))^2}{2\delta} I\{\frac{|X+Y-\mu_X(P)-\mu_Y(P)|}{2\delta} > \lambda\}\right] \ . \end{aligned}$$

Using the fact that $(A+B)^2 \leq 2(A^2 + B^2)$, we have further that this last expression is bounded above by

$$\begin{aligned} & E_P\left[\frac{(X-\mu_X(P))^2}{\delta} I\{\frac{|X-\mu_X(P)|}{2\delta} > \lambda\}\right] \\ + \ & E_P\left[\frac{(Y-\mu_Y(P))^2}{\delta} I\{\frac{|Y-\mu_Y(P)|}{2\delta} > \lambda\}\right] \ . \end{aligned}$$

Thus, by the assumption on the marginal distributions of $P \in \mathbf{P}$, we have that

$$\lim_{\lambda \to 0} \sup_{P \in \mathbf{P}} E_P\left[\frac{(X+Y-\mu_X(P)-\mu_Y(P))^2}{\sigma^2_{X+Y}(P)} I\{\frac{|X+Y-\mu_X(P)-\mu_Y(P)|}{\sigma_{X+Y}(P)} > \lambda\}\right] = 0 \ ,$$

which establishes the desired convergence. By Theorem 2.11 of Bhattacharya and Rao (1976), we have that

$$\sup_{S \in \mathcal{S}} |P_{n_j}\{(\sqrt{n_j}(\bar{X}_{n_j} - \mu_X(P_{n_j})), \sqrt{n_j}(\bar{Y}_{n_j} - \mu_Y(P_{n_j})))' \in S\} - \Phi_{1,1,\rho}(S)\}| \to 0 \ .$$

Since we also have that $\sup_{S \in \mathcal{S}} |\Phi(1, 1, \rho(P_{n_j}))(S) - \Phi_{1,1,\rho}(S)| \to 0$, this yields the desired contradiction in the first case.

Now consider the second case in which $|\rho| = 1$. Assume without loss of generality that $\rho = 1$. In this case, we have that the variance of $\sqrt{n}(\bar{X}_n - \bar{Y}_n)$ equals $2(1 - \rho(P_{n_j}))$, and therefore tends to 0. As a result,

$$\sqrt{n}(\bar{X}_n - \bar{Y}_n) \overset{P_{n_j}}{\to} 0 \ . \tag{38}$$

Consider $x < y$ and note that because of the convergence (38) and Lemma 11.4.1 of Lehmann and Romano (2005), we have that

$$P_{n_j}\{\sqrt{n_j}(\bar{X}_{n_j} - \mu_X(P_{n_j})) \leq x, \sqrt{n_j}(\bar{Y}_{n_j} - \mu_Y(P_{n_j})) \leq y\} \to \Phi(x) \ ,$$

where Φ is the standard normal distribution. Symmetrically, for $x > y$, we have that

$$P_{n_j}\{\sqrt{n_j}(\bar{X}_{n_j} - \mu_X(P_{n_j})) \leq x, \sqrt{n_j}(\bar{Y}_{n_j} - \mu_Y(P_{n_j})) \leq y\} \to \Phi(y) \ .$$

By applying these arguments to $y = x + \epsilon$ and $y = x - \epsilon$ for ϵ arbitrarily small, we have that

$$P_{n_j}\{\sqrt{n_j}(\bar{X}_{n_j} - \mu_X(P_{n_j})) \leq x, \sqrt{n_j}(\bar{Y}_{n_j} - \mu_Y(P_{n_j})) \leq x\} \to \Phi(x) \ .$$

Thus, by the Pormanteau Lemma (see, for example, van der Vaart (1998)), we have that

$$(\sqrt{n_j}(\bar{X}_{n_j} - \mu_X(P_{n_j})), \sqrt{n_j}(\bar{Y}_{n_j} - \mu_Y(P_{n_j})))' \overset{\mathcal{L}}{\to} \Phi_{1,1,1}$$

under P_{n_j}. Again, by Theorem 2.11 of Bhattacharya and Rao (1976), it follows that

$$\sup_{S \in \mathcal{S}} |P_{n_j}\{(\sqrt{n_j}(\bar{X}_{n_j} - \mu_X(P_{n_j})), \sqrt{n_j}(\bar{Y}_{n_j} - \mu_Y(P_{n_j})))' \in S\} - \Phi_{1,1,1}(S)\}| \to 0 \ ,$$

where here we are relying upon the assumption that $\Phi_{1,1,1}(\partial S) = 0$. Since we also have that $\sup_{S \in \mathcal{S}} |\Phi(1, 1, \rho(P_{n_j}))(S) - \Phi_{1,1,1}(S)| \to 0$, this yields the desired contradiction in the second case. ■

References

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

ANDREWS, D. W. K., BERRY, S. and JIA, P. (2004). Confidence regions for parameters in discrete games with multiple equilibria with an application to discount chain store location. Working paper, Department of Economics, Yale University.

BAHADUR, R. R. and SAVAGE, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics*, **27** 1115–1122.

BENKARD, C. L., BAJARI, P. and LEVIN, J. (2005). Estimating dynamic models of imperfect competition. Working paper, Graduate School of Business, Stanford University.

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

BORZEKOWSKI, R. and COHEN, A. (2005). Estimating strategic complementarities in credit unions' outsourcing decisions. Working paper, Federal Reserve Board of Governors.

CHERNOZHUKOV, V., HONG, H. and TAMER, E. (2004). Inference for identified parameter sets in econometric models. Working paper, Department of Economics, M.I.T.

CILBERTO, F. and TAMER, E. (2004). Market structure and multiple equilibria in airline markets. Working paper, Department of Economics, Northwestern.

DICICCIO, T. and ROMANO, J. P. (1990). Nonparametric confidence limits by resampling and least favorable distributions. *International Statistical Review*, **58** 59–76.

IMBENS, G. and MANSKI, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, **72** 1845–1857.

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*. 3rd ed. Springer, New York.

Manski, C. F. (2003). *Partial Identification of Probability Distributions.* Springer-Verlag, New York.

Manski, C. F. and Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, **70** 519–546.

Pakes, A., Porter, J., Ho, K. and Ishii, J. (2005). Moment inequalities and their application. Working paper, Department of Economics, Harvard University.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling.* Springer, New York.

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.

Romano, J. P. and Shaikh, A. M. (2006). Inference for the identified set in partially identified econometric models. Technical Report 2006–10, Department of Statistics, Stanford University.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.