Inference for the Identified Set 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 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 are 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 satisfying 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. Finally, we illustrate the use of our methods with an empirical study of the impact of top-coding of outcomes on inferences about the parameters of a linear regression.

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

1 Introduction

Recent empirical work in economics has shown that partially identified econometric models arise naturally in many contexts. By a partially identified model we mean 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 are 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

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

We henceforth refer to $\Theta_0(P)$ 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 the identified set with at least some prespecified probability asymptotically. That is, we require

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

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 $\mathcal{C}_n$ satisfying (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 $\mathcal{C}_n$ 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.

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 $X_i, i = 1, \dots, n$, generated from P, we seek random sets $\mathcal{C}_n = \mathcal{C}_n(X_1, \dots, X_n)$ 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 (2006).

To further motivate our study of partially identified models, we now provide a few concrete examples of such models. We will return to these examples later to provide illustrations of our methodology.

Example 1.1 (*One-Sided Mean*) Perhaps the simplest example of a partially identified model is given by the following setup. Let $X_i, i = 1, \dots, n$, 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 another distribution Q on $\mathbf{R}$ that is known to satisfy $\mu(Q) \geq \mu(P)$. 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 1.2 (*Two-Sided Mean*) A natural generalization of Example 1.1 is to consider bivariate random variables. To this end, let $(X_i, Y_i), i = 1, \dots, n$, 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 $\hat{Q}_n(\theta) = (\bar{X}_n - \theta)_+^2 + (\theta - \bar{Y}_n)_+^2$.

To provide an empirical context for this setup, consider the following special case. 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 the above setup. ■

Example 1.3 (*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.

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 $Y_i^* \leq 0$. 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 1.4 (*Moment Inequalities*) Consider the following generalization of Examples 1.1 and 1.2. 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

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

Of course, this set may be written 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.

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. ■

Our results on confidence regions for the identified set 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. 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 $\Theta_0(P)$. Moreover, our results provide a justification for iterating their procedure to produce ever smaller confidence regions until a stopping criterion is met 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, 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 the confidence regions in this paper have good uniform behavior. Concretely, we provide conditions under which $\mathcal{C}_n$ satisfies

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

By analogy with our earlier terminology, sets satisfying (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) 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. For this special case, they construct confidence regions that contain each identifiable parameter with at least some prespecified probability asymptotically and are valid uniformly in P. Romano and Shaikh (2006) construct 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,

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

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

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

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

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

By analogy with our earlier terminology, sets satisfying (6) are referred to as *confidence regions for a function of 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 satisfying these alternative coverage requirements.

Economists have used techniques for drawing inferences in partially identified models to solve many empirical problems that previously either were intractable or relied on untenable assumptions to achieve identification. In the context of missing outcome data and the analysis of treatment response, Manski (2003) has argued forcefully that reliance on such assumptions to make inferences is unfortunate, as it degrades the credibility of inferences made under those assumptions. A prominent example of this problem in other parts of empirical work in economics is provided by game-theoretic models that may have multiple equilibria. When confronted with such models, researchers often impose identification by assuming some sort of an ad hoc equilibrium selection mechanism. An alternative to this approach can be 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 can often be written as a system of moment inequalities, which can be viewed as a special case of the class of models we are considering. 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. 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 satisfying (5) and (6). We also briefly discuss an alternative, but less general approach to constructing confidence regions for the identified set. We provide an illustration of our methods in Section 3, wherein we study empirically the impact of top-coding of outcomes on inferences about the parameters in a linear regression. Finally, we conclude and provide directions for future research in Section 4.

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 satisfying (2) before turning our attention to the problem of constructing sets satisfying (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 satisfying (2) is equivalent to a certain multiple hypothesis testing problem. The problem is to test the family of hypotheses

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

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,

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

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

$$\limsup_{n \to \infty} FWER_P \leq \alpha \ . \tag{9}$$

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, $\mathcal{C}_n$, satisfies*

$$P\{\Theta_0(P) \subseteq \mathcal{C}_n\} = 1 - FWER_P \ ,$$

where $\Theta_0(P)$ is defined by (1). Conversely, given any random set $\mathcal{C}_n$, the procedure for testing the family of hypotheses (7) in which a null hypothesis H_θ is accepted iff $\theta \in \mathcal{C}_n$ satisfies

$$FWER_P = 1 - P\{\Theta_0(P) \subseteq \mathcal{C}_n\} \ .$$

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

$$\begin{aligned} P\{\Theta_0(P) \subseteq \mathcal{C}_n\} &= P\{\text{reject no null hypothesis } H_\theta \text{ s.t. } Q(\theta, P) = 0\} \\ &= 1 - FWER_P \ . \end{aligned}$$

The second conclusion follows from the same reasoning. ■

It follows from Lemma 2.1 that given any procedure for testing the family of null hypotheses (7) satisfying (9), the set of θ values corresponding to the set of accepted hypotheses, $\mathcal{C}_n$, satisfies (2). We thus turn our attention to analyzing 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.

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

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

under P; that is,

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

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 in order to determine its significance. Clearly, such a test satisfies $FWER_P \le \alpha$. To see this, note that

$$\begin{aligned} FWER_P &= P\{a_n\hat{Q}_n(\theta) > c_n(\Theta_0(P), 1-\alpha, P) \text{ for some } \theta \in \Theta_0(P)\} \\ &= 1 - P\{\sup_{\theta \in \Theta_0(P)} a_n\hat{Q}_n(\theta) \le c_n(\Theta_0(P), 1-\alpha, P)\} \le \alpha \ . \end{aligned}$$

But this test is infeasible, as the critical value depends on the unknown P. A crude solution to this difficulty is available if we are given estimates $\hat{c}_n(K, 1-\alpha)$ of the $1-\alpha$ quantile of the distribution of (10) that satisfy two properties. First, we require that the estimates $\hat{c}_n(K, 1-\alpha)$ be at least conservative under the joint null hypothesis $Q(\theta, P) = 0$ for all $\theta \in K$. Second, we require that the estimates are monotone in the sense that

$$\hat{c}_n(K, 1-\alpha) \ge \hat{c}_n(\Theta_0(P), 1-\alpha) \text{ for any } K \supseteq \Theta_0(P) \ . \tag{11}$$

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 estimate 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 of the $FWER_P$ at level α provided that these two assumptions are satisfied.

We will refrain from giving a more precise statement of the conditions under which such a test procedure provides asymptotic control of the $FWER_P$ because the results will follow from the analysis of the more powerful stepdown method in the next section.

2.1.3 Stepdown Control of the Familywise Error Rate

Stepdown 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 consider 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.
2. If $\sup_{\theta \in S_2} a_n\hat{Q}_n(\theta) \leq \hat{c}_n(S_2, 1-\alpha)$, then accept all hypotheses H_θ with $\theta \in S_2$ and stop; otherwise, set $S_3 = \{\theta \in \Theta : a_n\hat{Q}_n(\theta) \leq \hat{c}_n(S_2, 1-\alpha)\}$ 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

$$\limsup_{n \to \infty} P\{ \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) > \hat{c}_n(\Theta_0(P), 1-\alpha)\} \leq \alpha \ . \tag{12}$$

Theorem 2.1 *Let P denote the true distribution generating the data. Consider Algorithm 2.1 with critical values satisfying (11). Then,*

$$FWER_P \leq P\{ \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) > \hat{c}_n(\Theta_0(P), 1-\alpha)\} \ . \tag{13}$$

Hence, if the critical values also satisfy (12), then

$$\limsup_{n \to \infty} FWER_P \leq \alpha \ .$$

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 (11) 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. ■

Remark 2.1 For $S \subseteq \Theta$, let H_S denote the null hypothesis that $S = \Theta_0(P)$. The second condition (12) used in the proof of Theorem 2.1 essentially requires that the critical values $\hat{c}_n(S, 1-\alpha)$ asymptotically control the probability of a Type 1 error at level α when testing the null hypothesis H_S using the test statistic $\sup_{\theta \in S} a_n \hat{Q}_n(\theta)$. Thus, Theorem 2.1 remarkably reduces the multiple testing problem to the problem of constructing valid tests of each of the null hypotheses H_S. ■

Remark 2.2 It is worthwhile to note that the bound (13) is valid in finite samples. Thus, if critical values for testing the joint null hypotheses that $Q(\theta, P) = 0$ for all $\theta \in K$, $K \subseteq \Theta_0(P)$ that control the usual probability of a Type 1 error in finite samples are available, then Algorithm 2.1 with these critical values provides finite sample control of the $FWER_P$. Unfortunately, in the settings that are of interest to us, rarely will such critical values exist, and so we will have to resort to asymptotic arguments. ■

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.

For $K \subseteq \Theta$ and $\alpha \in (0, 1)$, define

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

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

Theorem 2.2 *Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P. 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 $J(\cdot, P)$ is continuous at its smallest $1 - \alpha$ quantile. Then, the following are true:*

(i) Condition (12) 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 et al. (1999). The second follows from Theorem 2.1. The third follows from Lemma 2.1. ■

Remark 2.3 As we will see below, in some cases, though it may be possible to assert that $J_n(\cdot, P)$ converges in distribution to a limit distribution $J(\cdot, P)$, it may be difficult to determine whether or not the limit distribution $J(\cdot, P)$ is continuous at its smallest $1 - \alpha$ quantile. In these instances, the conclusion of Theorem 2.2 will still hold if $\hat{c}_n(\Theta_0(P), 1 - \alpha)$ is given by $\hat{r}_n(K, 1 - \alpha) + \epsilon$ with $K = \Theta_0(P)$ and $\epsilon > 0$. A formal justification of this assertion is provided in the appendix. ■

Remark 2.4 Because $\binom{n}{b}$ may be large, it is often more practical to use the following approximation to (14). Let B_n be a positive sequence of numbers tending to ∞ as $n \to \infty$ and let $I_1, \dots, I_{B_n}$ be chosen randomly with or without replacement from the numbers $1, \dots, N_n$. Then, it follows from Corollary 2.4.1 of Politis et al. (1999) that one may approximate (14) by

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

without affecting the conclusions of Theorem 2.2. ■

We now revisit each of the examples described in the introduction and use Theorem 2.2 to provide conditions under which Algorithm 2.1 with $\hat{c}_n(\Theta_0(P), 1-\alpha)$ 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 (*One-Sided Mean*) Recall the setup of Example 1.1. Suppose P is such that the variance $\sigma^2(P)$ exists and is nonzero. It is straightforward to verify the assumptions of Theorem 2.2 for any such P. Let $a_n = n$. Assume without loss of generality that $\mu(P) = 0$. Then,

$$\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) = \sup_{\theta \ge 0} n(\bar{X}_n - \theta)^2_+ = (\sqrt{n}\bar{X}_n)^2_+ \xrightarrow{\mathcal{L}} (\sigma(P)Z)^2_+ \ ,$$

where Z is a standard normal random variable. This limit distribution is continuous at its $1 - \alpha$ quantile provided $\alpha < \frac{1}{2}$. ■

Example 2.2 (*Two-Sided Mean*) Recall the setup of Example 1.2. Suppose that P is such that the variances $\sigma^2_X(P)$ and $\sigma^2_Y(P)$ (and thus the covariance $\sigma_{X,Y}(P)$) exist and are both nonzero. Suppose further that $\mu_X(P) \leq \mu_Y(P)$. As in Example 2.1, it is straightforward to verify the assumptions of Theorem 2.2 for any such P. Let $a_n = n$. There are two cases to consider: either $\mu_X(P) < \mu_Y(P)$ or $\mu_X(P) = \mu_Y(P)$. In the former case, we have that with probability tending to 1, $\bar{X}_n < \bar{Y}_n$ eventually. Hence, only one of $(\bar{X}_n - \theta)^2_+$ or $(\theta - \bar{Y}_n)^2_+$ will be strictly positive. It is easy to see that the point of maximum of $\hat{Q}_n(\theta)$ over $\theta \in \Theta_0(P)$ will occur on the boundary. Thus, $\sup_{\theta \in \Theta_0(P)} n\hat{Q}_n(\theta)$ equals

$$\max\{n\hat{Q}_n(\mu_X(P)), n\hat{Q}_n(\mu_Y(P))\} \xrightarrow{\mathcal{L}} \max\{(\sigma_X(P)Z_1)^2_+, (\sigma_Y(P)Z_2)^2_+\}$$

where Z_1 and Z_2 are distributed as standard normal random variables with covariance $-\frac{\sigma_{X,Y}(P)}{\sigma_X(P)\sigma_Y(P)}$. This limit distribution is continuous at its $1-\alpha$ quantile provided $\alpha < \frac{1}{2}$. In the latter case, it is easy to see that

$$\sup_{\theta \in \Theta_0(P)} n\hat{Q}_n(\theta) = n\hat{Q}_n(\mu_1(P)) \xrightarrow{\mathcal{L}} (\sigma_1(P)Z_1)^2_+ + (\sigma_2(P)Z_2)^2_+ \ ,$$

where Z_1 and Z_2 are distributed as before. Again, this limit distribution is continuous at its $1-\alpha$ quantile provided $\alpha < \frac{1}{2}$. ∎

Example 2.3 (*Regression with Interval Outcomes*) Recall the setup of Example 1.3. Let $a_n = n$ and let $\{x_1, \ldots, x_J\}$ 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, $\sigma^2_1(P)$ 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

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

Chernozhukov et al. (2004) prove under these conditions that $J_n(\cdot, P)$ converges in distribution to a limit distribution $J(\cdot, P)$. Unfortunately, it is difficult to determine for which α the distribution of $\zeta(P)$ is continuous at its $1-\alpha$ quantile. Even so, Algorithm 2.1 may be used to construct confidence regions satisfying (2) following Remark 2.3. ∎

Example 2.4 (*Moment Inequalities*) Recall the setup of Example 1.4. Let $a_n = n$ and for $\delta \in \mathbf{R}^m_+$, let

$$K_n(\delta, P) = \{\theta \in \Theta : -\delta_j \leq \sqrt{n} E_P\{g_j(X_i, \theta)\} \leq 0 \text{ for all } j = 1, \ldots, m\} .$$

Suppose P is such that (i) for each $j = 1, \ldots, m$

$$\sup_{\theta \in \Theta_0(P)} \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (g_j(X_i, \theta) - E_P\{g_j(X_i, \theta)\}) = O_P(1) . \tag{15}$$

and (ii) for each $\delta \in \mathbf{R}^m_+$,

$$\sup_{\theta \in K_n(\delta, P)} \sum_{1 \leq j \leq m} (\frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (g_j(X_i, \theta))^2_+ \xrightarrow{\mathcal{L}} \zeta(\delta, P) .$$

With an abuse of notation, let

$$K_n(\delta, P)^{\mathrm{c}} = \{\theta \in \Theta_0(P) : \theta \notin K_n(\delta, P)\} .$$

Using this notation, we have that $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta)$ may be written as

$$\sup_{\theta \in K_n(\delta, P)} a_n \hat{Q}_n(\theta) \vee \sup_{\theta \in K_n(\delta, P)^{\mathrm{c}}} a_n \hat{Q}_n(\theta) .$$

Note that the first of these two terms, $\sup_{\theta \in K_n(\delta, P)} a_n \hat{Q}_n(\theta)$, equals

$$\sup_{\theta \in K_n(\delta, P)^{\mathrm{c}}} \sum_{1 \leq j \leq m} \{\frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (g_j(X_i, \theta) - E_P\{g_j(X_i, \theta)\} + \sqrt{n} E_P\{g_j(X_i, \theta)\})^2_+\} .$$

Using the defintion of $K_n(\delta, P)$ and (15), it is not difficult to see that this expression equals 0 with probability tending to 1 as n and δ both tend to infinity. By assumption, we have that the second term, $\sup_{\theta \in K_n(\delta, P)} a_n \hat{Q}_n(\theta)$, converges in distribution to $\zeta(\delta, P)$. Define $\zeta(P) = \lim_{\delta \to \infty} \zeta(\delta, P)$. Note that this limit exists by the Monotone Convergence Theorem and is tight because $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta)$ is bounded above by

$$\sup_{\theta \in \Theta_0(P)} \sum_{1 \leq j \leq m} \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (g_j(X_i, \theta) - E_P\{g_j(X_i, \theta)\})^2_+ .$$

Following the arguments given in Chernozhukov et al. (2004), we may now assert that

$$\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \xrightarrow{\mathcal{L}} \zeta(P) .$$

Unfortunately, as in Example 2.3, it is difficult to determine for which α the distribution of $\zeta(P)$ is continuous at its $1 - \alpha$ quantile, but Algorithm 2.1 may still be used to construct confidence regions satisfying (2) following Remark 2.3. ■

Remark 2.5 Throughout this paper, we are assuming 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

$$\hat{r}_n(K, 1-\alpha) = \inf\{x : \frac{1}{n-b+1} \sum_{1 \leq i \leq n-b+1} I\{\sup_{\theta \in K} a_b \hat{Q}_{n,b,i}(\theta) \leq x\} \geq 1-\alpha\} \ , \qquad (16)$$

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 et al. (1999) that the conclusions of the theorem continue to hold. ■

Remark 2.6 Our construction of critical values has used subsampling. We now show by example that the bootstrap may fail to approximate the distribution of the statistic

$$\sup_{\theta \in K} a_n \hat{Q}_n(\theta) = (\sqrt{n}\bar{X}_n)_+^2 \qquad (17)$$

when $K = \Theta_0(P)$. Note that $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \sim (Z)_+^2$, where Z is a standard normal random variable.

To this end, recall the setup of Example 1.1. Let $a_n = n$ and suppose that $P = N(0,1)$. 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

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

First note that $P\{A_c\} = 1$ for any $c \in (0, \infty)$, where

$$A_c = \{\omega : \liminf_{n \to \infty} \sqrt{n}\bar{X}_n < -c\} \ . \qquad (18)$$

To see this, note that 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 (17) at $K = \Theta_0(P) = [0, \infty)$, which is given by

$$\mathcal{L}(\sup_{\theta \in \Theta_0(P)} a_n(\hat{Q}_n^*(\theta) - \hat{Q}_n(\theta)) | X_i, i = 1, \ldots, n) \ . \tag{19}$$

This approximation mimics the hypothesis that $Q(\theta, P) = 0$ for all $\theta \in \Theta_0(P)$ by centering about $\hat{Q}_n(\theta)$. 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

$$\begin{aligned}
\sup_{\theta \in \Theta_0(P)} a_{n_k}(\hat{Q}^*_{n_k}(\theta) - \hat{Q}_{n_k}(\theta)(\omega)) &= \sup_{\theta \geq 0}(n_k(\bar{X}^*_{n_k} - \theta)^2_+ - n_k(\bar{X}_{n_k}(\omega) - \theta)^2_+) \\
&= \sup_{\theta \geq 0} n_k(\bar{X}^*_{n_k} - \theta)^2_+ \\
&= (\sqrt{n_k}\bar{X}^*_{n_k})^2_+ \\
&= (\sqrt{n_k}(\bar{X}^*_{n_k} - \bar{X}_{n_k}(\omega)) + \sqrt{n_k}\bar{X}_{n_k}(\omega))^2_+ \\
&\leq (\sqrt{n_k}(\bar{X}^*_{n_k} - \bar{X}_{n_k}(\omega)) - c)^2_+ \\
&\stackrel{\mathcal{L}}{\to} (Z - c)^2_+ \ .
\end{aligned}$$

The second equality follows from the fact that $\sqrt{n_k}(\bar{X}_{n_k}(\omega) - \theta) < 0$ for all $\theta \geq 0$ since $\sqrt{n_k}\bar{X}_{n_k}(\omega) < -c$. It follows that the bootstrap fails to approximate the distribution of (17) at $K = \Theta_0(P)$.

It is worthwhile to consider the actual value of the probability (12) if the bootstrap approximation above were used instead of subsampling. To this end, first note that this probability is given by

$$P\{\sup_{\theta \in \Theta_0(P)} a_n\hat{Q}_n(\theta) > \hat{c}_{\text{boot},n}(1 - \alpha)\} \ , \tag{20}$$

where $\hat{c}_{\text{boot},n}(1 - \alpha)$ is the $1 - \alpha$ quantile of (19). Recall that $\sup_{\theta \in \Theta_0(P)} a_n\hat{Q}_n(\theta)$ is distributed as $(Z)^2_+$, where Z is a standard normal random variable. Note that

$\sup_{\theta \in \Theta_0(P)} a_n(\hat{Q}_n^*(\theta) - \hat{Q}_n(\theta))$ equals

$$\begin{aligned} & \sup_{\theta \in \Theta_0(P)} n((\bar{X}_n^* - \theta)_+^2 - (\bar{X}_n - \theta)_+^2) \\ = \; & \sup_{\theta \geq 0} (\sqrt{n}(\bar{X}_n^* - \bar{X}_n) + \sqrt{n}\bar{X}_n - \sqrt{n}\theta)_+^2 - (\sqrt{n}\bar{X}_n - \sqrt{n}\theta)_+^2) \; . \end{aligned}$$

Therefore, (19) converges in distribution to

$$\mathcal{L}(\sup_{\theta \geq 0} (Z^* + Z - \theta)_+^2 - (Z - \theta)_+^2 | Z) \; , \tag{21}$$

where Z^* is a standard normal random variable distributed independently of Z. The probability (20) is therefore asymptotically equal to

$$P\{(Z)_+^2 > c(1 - \alpha | Z)\} \; , \tag{22}$$

where $c(1 - \alpha|Z)$ is the $1 - \alpha$ quantile of (21). Note that (21) dominates

$$\mathcal{L}((Z^* + Z)_+^2 - (Z)_+^2 | Z) \; . \tag{23}$$

It follows that $c(1 - \alpha|Z)$ is no smaller than $c'(1 - \alpha|Z)$, the $1 - \alpha$ quantile of (23). The probability (22) is therefore bounded above by

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

Using this bound, we may use simulation to determine that (22) is asymptotically bounded above by approximately 0 for $\alpha = .1$ and $\alpha = .05$. In fact, with 2000 simulations we find that (22) is asymptotically bounded above by 0 to three significant digits for both $\alpha = .1$ and $\alpha = .05$. Therefore, while critical values from the bootstrap satisfy (12), they are, as before, too conservative for practical purposes. ■

Remark 2.7 It is interesting to note that the proof of Theorem 2.1 only requires that the initial set S_1 be such that $\Theta_0(P) \subseteq S_1$. Indeed, one can allow for S_1 to be random provided that it satisfies

$$P\{\Theta_0(P) \subseteq S_1\} \to 1 \tag{24}$$

without affecting the argument in any way. Thus, using our results, it is possible to recast the method of Chernozhukov et al. (2004) as a single-step method with a particular choice of S_1. In their procedure,

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

where ϵ_n is a positive sequence of constants tending to zero slowly. Recall that because of this rate restriction on ϵ_n, they are able to show that (24) holds. Hence, using our results, it follows that their confidence regions satisfy (2). Unfortunately, the level of ϵ_n is completely 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, thereby removing this arbitrariness, and produce ever smaller confidence regions until a stopping criterion is met while still maintaining the coverage requirement.

The authors' goal, however, is not confidence regions that satisfy (2), but instead ones that satisfy the stronger coverage property

$$\lim_{n\to\infty} P\{\Theta_0(P) \subseteq \mathcal{C}_n\} = 1 - \alpha \ . \tag{25}$$

They show that their confidence regions satisfy (25) under certain assumptions, including the assumptions of our Theorem 2.2, and verify that these assumptions hold for several examples. Consider the confidence region resulting from our stepdown procedure with $S_1 = \hat{\Theta}_{0,n}$. Since confidence regions produced in this way are smaller than theirs, but still satisfies the level constraint, it follows that such confidence regions satisfy (25) whenever their assumptions hold. ■

Remark 2.8 It follows from the discussion in Remark 2.7 that there are no first-order asymptotic differences between confidence regions from our stepdown procedure with $S_1 = \hat{\Theta}_{0,n}$ and those of Chernozhukov et al. (2004). 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. Consider Example 1.1. Let $a_n = n$ and suppose P is such that $\sigma^2(P)$ exists. Assume further that $E_P\{X_i^4\}$ exists and that Cramér's condition holds; that is,

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

where $\psi_P(s)$ denotes the characteristic function of P. Let

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

where $\lambda_n > 0$ is an increasing sequence tending to infinity, but so slowly that $\lambda_n/n \to 0$. These rate restrictions on λ_n ensure that the assumptions of Chernozhukov et al. (2004) are satisfied. The confidence region of Chernozhukov et al. (2004) is given by

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

In order to obtain a second-order accurate expression for $\hat{r}_n(\hat{\Theta}_{0,n}, 1-\alpha)$, first note that $\hat{r}_n(\hat{\Theta}_{0,n}, 1-\alpha)$ is the $1-\alpha$ quantile of

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

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

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

Now consider

$$\tilde{L}_n(x) = \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\{\sqrt{b}(\bar{X}_{n,b,i} - \bar{X}_n) \leq x\} \ .$$

Since $\sqrt{b}(\bar{X}_n - \mu(P)) = O_P(\sqrt{b/n})$, it follows from Lemma 5.2 of Romano and Shaikh (2006) that

$$\tilde{L}_n(x) = J_b(x, P) + O_P(\sqrt{\frac{b}{n}})$$

uniformly in x, where $J_b(x, P) = P\{\sqrt{b}(\bar{X}_b - \mu(P)) \leq x\}$. From Theorem 15.5.1 of Lehmann and Romano (2005), we have that

$$J_b(x, P) = \Phi(\frac{x}{\sigma(P)}) - \frac{1}{6\sqrt{b}}\gamma(P)\phi(\frac{x}{\sigma(P)})(\frac{x^2}{\sigma^2(P)} - 1) + O(\frac{1}{b})$$

uniformly in x, where

$$\gamma(P) = E_P\{(\frac{X_i - \mu(P)}{\sigma(P)})^3\} \ .$$

It follows that

$$\tilde{L}_n(x) = \Phi(\frac{x}{\sigma(P)}) - \frac{1}{6\sqrt{b}}\gamma(P)\phi(\frac{x}{\sigma(P)})(\frac{x^2}{\sigma^2(P)} - 1) + O_P(\frac{1}{b} \vee \sqrt{\frac{b}{n}})$$

uniformly in x. From this we can deduce that the $1-\alpha$ quantile of $\tilde{L}_n(x)$ is

$$\sigma(P)(z_{1-\alpha} + \frac{\delta}{\sqrt{b}}) + O_P(\frac{1}{b} \vee \sqrt{\frac{b}{n}}) \ ,$$

where $\delta = \frac{1}{6}\gamma(P)(z_{1-\alpha}^2 - 1)$ and $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. Hence,

$$\hat{r}_n(\hat{\Theta}_{0,n}, 1-\alpha) = (\sigma(P)(z_{1-\alpha} + \frac{\delta}{\sqrt{b}}) + \sqrt{\frac{\lambda_n b}{n}} + O_P(\frac{1}{b} \vee \sqrt{\frac{b}{n}}))^2 .$$

Now consider the confidence region $\mathcal{C}_n$ given by Algorithm 2.1. First note that $\sup_{\theta \in \hat{\Theta}_{0,n}} a_n \hat{Q}_n(\theta) = \lambda_n^2$. From the above expression for $\hat{r}_n(\hat{\Theta}_{0,n}, 1-\alpha)$, we therefore have that

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

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

$$\{\theta \in \mathbf{R} : n(\bar{X}_n - \theta) \leq \hat{r}_n(\mathcal{C}'_n, 1-\alpha)\}$$

with probability tending to 1. From the above analysis, we have immediately that $\hat{r}_n(\mathcal{C}'_n, 1-\alpha)$ is given by

$$(\sigma(P)(z_{1-\alpha} + \frac{\delta}{\sqrt{b}}) + \sqrt{\frac{b}{n}}(\sigma(P)(z_{1-\alpha} + \frac{\delta}{\sqrt{b}}) + \sqrt{\frac{\lambda_n b}{n}}) + O_P(\frac{1}{b} \vee \sqrt{\frac{b}{n}}))^2 .$$

To complete the comparison, suppose first that $1/b$ is of order no larger than $\sqrt{b/n}$; that is, $n/b^3 = O(1)$. If this is the case, then it follows from the above expressions for $\hat{r}_n(\hat{\Theta}_{0,n}, 1-\alpha)$ and $\hat{r}_n(\mathcal{C}'_n, 1-\alpha)$ that $\mathcal{C}_n$ is smaller to second-order than $\mathcal{C}'_n$. To see this, note that if $1/b$ is of order no larger than $\sqrt{b/n}$, then

$$\hat{r}_n(\hat{\Theta}_{0,n}, 1-\alpha) = (\sigma(P)(z_{1-\alpha} + \frac{\delta}{\sqrt{b}}) + O_P(\sqrt{\frac{\lambda_n b}{n}}))^2 ,$$

whereas

$$\hat{r}_n(\mathcal{C}'_n, 1-\alpha) = (\sigma(P)(z_{1-\alpha} + \frac{\delta}{\sqrt{b}}) + O_P(\sqrt{\frac{b}{n}}(1 \vee \sqrt{\frac{\lambda_n b}{n}})))^2 .$$

When $1/b$ is of order larger than $\sqrt{b/n}$, the comparison of $\mathcal{C}_n$ to $\mathcal{C}'_n$ becomes more delicate and depends on the rate at which λ_n tends to infinity, but by construction $\mathcal{C}_n$ can, of course, never be larger than $\mathcal{C}'_n$.

Finally, we note briefly that it is possible to to construct confidence regions that are even smaller to second-order than $\mathcal{C}_n$ above by considering test statistics in which the mean has been studentized; that is, replace the test statistics above with ones based on

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

where $\hat{\sigma}_n$ is the usual estimate of $\sigma(P)$. See Chapter 10 of Politis et al. (1999) for details. ■

2.1.5 An Alternative Construction

As an alternative to the construction of confidence regions satisfying (2) presented above, consider a single-step method in which

$$S_1 = \tilde{\Theta}_{0,n} = \{\theta \in \Theta : \hat{Q}_n(\theta) = 0\}$$

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

$$\mathcal{C}_n = \{\theta \in \Theta : a_n \hat{Q}_n(\theta) \leq \hat{r}_n(\tilde{\Theta}_{0,n}, 1-\alpha)\} \ . \tag{26}$$

Under the assumptions of Theorem 2.2, such an approach may often lead to confidence regions satisfying (2), as in the following two examples.

Example 2.5 (*One-Sided Mean*) Recall the setup of Example 1.1. Let $a_n = n$ and suppose P is such that the variance $\sigma^2(P)$ exists and is nonzero. From the analysis of Example 2.1, we have that $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \xrightarrow{\mathcal{L}} (\sigma(P)Z)_+^2$, where Z is a standard normal random variable. Note that this limit distribution is continuous at its $1-\alpha$ quantile provided $\alpha < \frac{1}{2}$. By Theroem 2.1.1 of Politis et al. (1999), to show that $\mathcal{C}_n$ defined by (26) satisfies (2), it therefore suffices to show that $\sup_{\theta \in \tilde{\Theta}_{0,n}} a_b \hat{Q}_b(\theta)$ shares this limit distribution. To this end, first note that $\tilde{\Theta}_{0,n} = \{\theta \in \Theta : \theta \geq \bar{X}_n\}$. Thus,

$$\sup_{\theta \in \tilde{\Theta}_{0,n}} a_b \hat{Q}_b(\theta) = b(\bar{X}_b - \bar{X}_n)_+^2 \xrightarrow{\mathcal{L}} (\sigma(P)Z)_+^2 \ ,$$

where Z is again a standard normal random variable. The desired conclusion follows. ■

Example 2.6 (*Two-Sided Mean*) Recall the setup of Example 1.2. Let $a_n = n$ and suppose that P is such that the variances $\sigma_X^2(P)$ and $\sigma_Y^2(P)$ (and thus the covariance $\sigma_{X,Y}(P)$) exist and are both nonzero. Suppose further that $\mu_X(P) \leq \mu_Y(P)$. From the analysis of Example 2.2, we have that

$$\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \xrightarrow{\mathcal{L}} \begin{cases} \max\{(\sigma_X(P)Z_1)_+^2, (\sigma_Y(P)Z_2)_+^2\} & \text{if } \mu_X(P) < \mu_Y(P) \\ (\sigma_1(P)Z_1)_+^2 + (\sigma_2(P)Z_2)_+^2 & \text{if } \mu_X(P) = \mu_Y(P) \end{cases} ,$$

where Z_1 and Z_2 are standard normal random variables with covariance $-\frac{\sigma_{X,Y}(P)}{\sigma_X(P)\sigma_Y(P)}$. Note that this limit distribution is continuous at its $1-\alpha$ quantile provided $\alpha < \frac{1}{2}$. Therefore, as in Example 2.5, to show that $\mathcal{C}_n$ defined by (26) satisfies (2), it suffices to show that $\sup_{\theta \in \tilde{\Theta}_{0,n}} a_b \hat{Q}_b(\theta)$ shares this limit distribution.

To this end, first consider the case in which $\mu_X(P) < \mu_Y(P)$. Then, with probability approaching 1, we have that $\bar{X}_n < \bar{Y}_n$, so

$$\tilde{\Theta}_{0,n} = \{\theta \in \mathbf{R} : \bar{X}_n < \theta < \bar{Y}_n\} \ .$$

Similarly, we also have that $\bar{X}_n < \bar{Y}_b$ and $\bar{X}_b < \bar{Y}_n$ with probability approaching 1. Thus,

$$\begin{aligned} \sup_{\theta \in \tilde{\Theta}_{0,n}} a_b \hat{Q}_b(\theta) &= \max\{a_b \hat{Q}_b(\bar{X}_n), a_b \hat{Q}_b(\bar{Y}_n)\} \\ &= \max\{b(\bar{X}_b - \bar{X}_n)^2_+ + b(\bar{X}_n - \bar{Y}_b)^2_+, b(\bar{X}_b - \bar{Y}_n)^2_+ + b(\bar{Y}_n - \bar{Y}_b)^2_+\} \\ &\stackrel{\mathcal{L}}{\to} \max\{(\sigma_X(P) Z_1)^2_+, (\sigma_Y(P) Z_2)^2_+\} \ . \end{aligned}$$

The desired conclusion now follows for this case.

Now consider the case in which $\mu_X(P) = \mu_Y(P)$. Then,

$$\hat{\Theta}_{0,n} = \begin{cases} \{\theta \in \mathbf{R} : \bar{X}_n < \theta < \bar{Y}_n\} & \text{if } \bar{X}_n < \bar{Y}_n \\ \{\frac{\bar{X}_n + \bar{Y}_n}{2}\} & \text{otherwise} \end{cases} .$$

Thus, if $\bar{X}_n < \bar{Y}_n$,

$$\begin{aligned} \sup_{\theta \in \tilde{\Theta}_{0,n}} a_b \hat{Q}_b(\theta) &= \max\{a_b \hat{Q}_b(\bar{X}_n), a_b \hat{Q}_b(\bar{Y}_n)\} \\ &= \max\{b(\bar{X}_b - \bar{X}_n)^2_+ + b(\bar{X}_n - \bar{Y}_b)^2_+, b(\bar{X}_b - \bar{Y}_n)^2_+ + b(\bar{Y}_n - \bar{Y}_b)^2_+\} \\ &\stackrel{\mathcal{L}}{\to} (\sigma_1(P) Z_1)^2_+ + (\sigma_2(P) Z_2)^2_+ \ . \end{aligned}$$

On the other hand, if $\bar{X}_n \geq \bar{Y}_n$,

$$\begin{aligned} \sup_{\theta \in \tilde{\Theta}_{0,n}} a_b \hat{Q}_b(\theta) &= a_b \hat{Q}_b(\frac{\bar{X}_n + \bar{Y}_n}{2}) \\ &= b(\bar{X}_b - \frac{\bar{X}_n + \bar{Y}_n}{2})^2_+ + b(\frac{\bar{X}_n + \bar{Y}_n}{2} - \bar{Y}_b)^2_+ \\ &\stackrel{\mathcal{L}}{\to} (\sigma_1(P) Z_1)^2_+ + (\sigma_2(P) Z_2)^2_+ \ . \end{aligned}$$

The desired conclusion now follows. ■

The following example, however, shows that the alternative construction given by (26) may not always produce valid confidence regions.

Example 2.7 Let $X_i, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution P. The identified set, $\Theta_0(P)$, is assumed to be the set of medians of P. For ease of argument, suppose it is known to the researcher that 0 is a median of P, but it may not be the only median. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = E_P\{|X - \theta| - |X|\} \ .$$

The sample analog of $Q(\theta, P)$ is simply $\frac{1}{n}\sum_{1\le i\le n} |X_i - \theta| - |X_i|$.

Let $a_n = \sqrt{n}$ and suppose P is defined by $P\{X_i = 1\} = P\{X_i = -1\} = \frac{1}{2}$. Hence, $\Theta_0(P) = [-1, 1]$. Note that

$$|x - \theta| - |x| = D(x)\theta + R(x, \theta) \ ,$$

where $D(x) = -I\{x > 0\} + I\{x \ge 0\}$ and $R(x, \theta) = 2(\theta - x)\text{sgn}(x)I\{|x| \le |\theta|\}$. Since $P\{|X| \le \theta\} = 0$ for any $\theta \in \Theta_0(P)$, it follows that with probability 1

$$\sup_{\theta\in\Theta_0(P)} a_n\hat{Q}_n(\theta) = \sup_{-1\le\theta\le 1} \frac{1}{\sqrt{n}} \sum_{1\le i\le n} D(X_i)\theta = \hat{Z}_n I\{\hat{Z}_n > 0\} - \hat{Z}_n I\{\hat{Z}_n \le 0\} \ ,$$

where $\hat{Z}_n = \frac{1}{\sqrt{n}}\sum_{1\le i\le n} D(X_i)$. Therefore, by the Continuous Mapping Theorem

$$\sup_{\theta\in\Theta_0(P)} a_n\hat{Q}_n(\theta) \xrightarrow{\mathcal{L}} ZI\{Z > 0\} - ZI\{Z \le 0\}) \ , \tag{27}$$

where Z is a standard normal random variable.

We will now show that asymptotically the distribution of $\sup_{\theta\in\tilde{\Theta}_{0,n}} a_b\hat{Q}_b(\theta)$ is not close to the distribution of $\sup_{\theta\in\Theta_0(P)} \sqrt{n}\hat{Q}_n(\theta)$. To this end, let n tend to infinity through the odd integers . As a result, $\tilde{\Theta}_{0,n}$ will always be a singleton. Let $\tilde{\theta}_n$ denote this point. It is easy to see that

$$\tilde{\theta}_n = \begin{cases} 1 & \text{if } \hat{Z}_n > 0 \\ -1 & \text{if } \hat{Z}_n < 0 \end{cases} \ . \tag{28}$$

Since $\tilde{\theta}_n \in \Theta_0(P)$, we have as before that $P\{|X| \leq \tilde{\theta}_n\} = 0$. It follows that with probability 1

$$\sup_{\theta \in \tilde{\Theta}_{0,n}} a_b \hat{Q}_b(\theta) = \frac{1}{\sqrt{b}} \sum_{1 \leq i \leq b} D(X_i)\tilde{\theta}_n = \hat{Z}_b I\{\hat{Z}_n > 0\} - \hat{Z}_b I\{\hat{Z}_n \leq 0\} \ .$$

Since $b/n \to 0$, $(\hat{Z}_b, \hat{Z}_n)'$ converges in distribution to a bivariate normal distribution with variances equal to 1 and correlation 0. Hence, by the Continuous Mapping Theorem, the distribution of $\sup_{\theta \in \tilde{\Theta}_{0,n}} a_B \hat{Q}_b(\theta)$ is asymptotically quite different from the distribution (27). Importantly, it's $1-\alpha$ quantile will asymptotically be strictly smaller than the $1-\alpha$ quantile of the distribution of (27). Thus, $\mathcal{C}_n$ defined by (26) fails to satisfy (2) for this example. ■

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. 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*

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

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

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

By Theorem 3.1(iv) of Romano and Shaikh (2006), it follows that

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

Thus,

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}} FWER_P \leq \alpha \ .$$

The asserted claim now follows immediatly from Lemma 2.1. ■

We now provide two applications of Theorem 2.3 to constructing confidence regions satisfying the coverage requirement (3).

Example 2.8 (*One-Sided Mean*) Recall the setup of Example 1.1. 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 $a_n = n$ and let $\mathbf{P}$ be a set of distributions satisfying

$$\lim_{\lambda\to\infty} \sup_{P\in\mathbf{P}} E_P\left[\frac{|X-\mu(P)|^2}{\sigma^2(P)} I\left\{\frac{|X-\mu(P)|}{\sigma(P)} > \lambda\right\}\right] = 0 \ . \tag{30}$$

Note that because $\sup_{\theta\in\Theta_0(P)} a_n\hat{Q}_n(\theta) \geq 0$, we may restrict attention to $x \geq 0$ in (29). Note further that for $x \geq 0$,

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

By Lemma 11.4.1 of Lehmann and Romano (2005), we have that

$$\sup_{P\in\mathbf{P}} \sup_{x\geq 0} |J_n(x, P) - \Phi_{\sigma(P)}(x)| \to 0 \ ,$$

where $\Phi_{\sigma(P)}$ is the distribution of a mean-zero normal random variable with variance $\sigma^2(P)$. The desired condition (29) follows from an appeal to the triangle inequality. ■

Example 2.9 (*Two-Sided Mean*) Recall the setup of Example 1.2. 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 a set of bivariate distributions such that the marginal distributions satisfy (30) and $X_i \leq Y_i$ with probability 1 (thus, $\mu_X(P) \leq \mu_Y(P)$). We will argue by contradiction that the required condition (29) holds. 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 $\mu_X(P_{n_j}) \leq \mu_Y(P_{n_j})$ and

$$\sup_{x\in\mathbf{R}}\{J_{b_{n_j}}(x, P_{n_j}) - J_{n_j}(x, P_{n_j})\} \to \delta \ . \tag{31}$$

for some $\delta > 0$. Note that since $\bar{X}_{b_{n_j}} < \bar{Y}_{b_{n_j}}$, then $\sup_{\theta \in \Theta_0(P_{n_j})} a_{b_{n_j}} \hat{Q}_{b_{n_j}}(\theta)$ equals

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

where

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

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

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

Since $\Delta_{b_{n_j}}(P_{n_j}) \geq \Delta_{n_j}(P_{n_j})$, we have that

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

Thus, (31) implies that

$$\sup_{x \in \mathbf{R}} \{\tilde{J}_{b_{n_j}}(x, P_{n_j}) - J_{n_j}(x, P_{n_j})\} \not\to 0 . \tag{32}$$

Since $a_n \hat{Q}_n(\theta) \geq 0$, we may restrict attention to $x \geq 0$. Next, note that for $x \geq 0$, $\tilde{J}_{b_{n_j}}(x, P_{n_j})$ is the probability under P_{n_j} that $(Z_{1,b_{n_j}}(P_{n_j}), Z_{2,b_{n_j}}(P_{n_j}))'$ lies in a set $S_x \in \mathcal{S}$, where

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

Importantly, $J_{n_j}(x, P_{n_j})$ is simply the probability under P_{n_j} that $(Z_{1,n_j}(P_{n_j}), Z_{2,n_j}(P_{n_j}))'$ lies in the same set. But, by Lemma 3.1 of Romano and Shaikh (2006), 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 (32). ■

2.3 Confidence Regions for Functions of the Identified Set

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

$$H_\lambda : \lambda \in \Lambda_0(P) \text{ for } \lambda \in \Lambda \ , \tag{33}$$

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

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

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

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

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

$$\inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \ , \tag{34}$$

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 our earlier test statistic $a_n \hat{Q}_n(\theta)$. For $K \subseteq \Lambda$, let $\hat{c}_n^f(K, 1-\alpha)$ be an estimate 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 $\sup_{\lambda \in S_1} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}^f_n(S_1, 1-\alpha)$, then accept all hypotheses and stop; otherwise, set $S_2 = \{\lambda \in \Lambda : \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}^f_n(S_1, 1-\alpha)\}$ and continue.

2. If $\sup_{\lambda \in S_2} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}^f_n(S_2, 1-\alpha)$, then accept all hypotheses H_λ with $\lambda \in S_2$ and stop; otherwise, set $S_3 = \{\lambda \in \Lambda : \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}^f_n(S_2, 1-\alpha)\}$ and continue.

   $\vdots$

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

   $\vdots$

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

$$\hat{r}^f_n(K, 1-\alpha) = \inf\{x : \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\{\sup_{\lambda \in K} \inf_{\theta \in f^{-1}(\lambda)} a_b \hat{Q}_{n,b,i}(\theta) \leq x\} \geq 1-\alpha\} \ . \tag{35}$$

Theorem 2.4 *Let $X_i, i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution P. 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 $\mathcal{C}^f_n$ denote the set of θ values corresponding to accepted hypotheses from Algorithm 2.2 when $\hat{c}^f_n(K, 1-\alpha)$ is given by (35).*

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

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

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}} \sup_{x \in \mathbf{R}} \{J_b(x, P) - J_n(x, P)\} \leq 0 \ . \tag{36}$$

Then, $\mathcal{C}^f_n$ satisfies (6).

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 to construct sets satisfying (5) and (6).

Example 2.10 Consider the following straightforward generalization of Example 1.1. 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 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 satisfying (5). To this end, let $a_n = n$ and suppose P is such that $\sigma^2_X(P)$ exists. Assume without loss of generality that $\mu_X(P) = 0$. Then,

$$\begin{aligned} \sup_{\theta_1 \in \Lambda_0(P)} \inf_{f^{-1}(\theta_1)} a_n \hat{Q}_n(\theta) &= \sup_{\theta_1 \geq 0} \inf_{\theta_2 \in \mathbf{R}} n(\bar{X}_n - \theta_1)^2_+ + n(\bar{Y}_n - \theta_2)^2_+ \\ &= \sup_{\theta_1 \geq 0} n(\bar{X}_n - \theta_1)^2_+ \\ &= n(\bar{X}_n)^2_+ \\ &\stackrel{\mathcal{L}}{\to} (\sigma_X(P) Z)^2_+ \ , \end{aligned}$$

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 $\hat{c}^f_n(K, 1-\alpha)$ is given by (35), $\mathcal{C}^f_n$, 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 (30). Since $\sup_{\theta_1 \in \Lambda_0(P)} \inf_{f^{-1}(\theta_1)} a_n \hat{Q}_n(\theta)$ is simply $n(\bar{X}_n)^2_+$, it follows immediately from the analysis of Example 2.8 that (36). 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 (35), $\mathcal{C}_n^f$, satisfies (6). ■

Remark 2.9 Of course, given a confidence region for the identified set $\mathcal{C}_n$, one crude construction of a confidence region for a function of the identified set is available as the image of $\mathcal{C}_n$ under the function of interest. Unfortunately, such a construction will typically be very conservative. ■

3 Empirical Illustration

In this section, we use the techniques developed above to examine the impact of top-coding of outcomes on the inferences that can be made about the parameters of a linear regression. By top-coding of 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 1.3, and so the theory developed above applies here under the appropriate assumptions.

The motivation for our exercise stems from the following observation. In order to study changes in the wage structure and earnings inequality, researchers often regress the logarithm of hourly wages on various demographic characteristics. Data sets that are used for this purpose invariably top-code wages for reasons of confidentiality. One approach to deal with the top-coding of wages is to simply 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. For a detailed description of such studies and documentation of this practice, see 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.

In order 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) are at least college graduates; (4) primary source of income is wages and salaries; and (5) worked at least 2 hours per week on average. There are 305 such observations, none of which suffer 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 this 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 1.3. In order 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 $Y_i^* = \log(\texttt{wage}_i^*/H_i)$, where $\texttt{wage}_i^*$ 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 $\texttt{wage}_i^*$ is bounded above by $\overline{\texttt{wage}} = \$10^8$. 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 $\underline{\texttt{wage}}$ be the threshhold above which wages are not observed. Define $Y_{1,i} = Y_{2,i} = Y_i^*$ if $\texttt{wage}_i^* \leq \underline{\texttt{wage}}$; otherwise, let $Y_{1,i} = \underline{Y}_i = \log(\underline{\texttt{wage}}/H_i)$ and $Y_{2,i} = \overline{Y}_i = \log(\overline{\texttt{wage}}/H_i)$.

Below we will construct both confidence regions 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 discussed in Romano and Shaikh (2006). We will compare the inferences that can be drawn from these confidence regions with the ones 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 $\texttt{wage}_i^* \leq \underline{\texttt{wage}}$ and $Y_i^a = 1.5 \times \underline{Y}_i$ otherwise.

Before proceeding, we discuss some computational details. First, consider the choice of the subsample size b. In practice, one would, of course, like to use a data-dependent subsample size; see Politis et al. (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 drive by variation in the choice of subsample size. Note

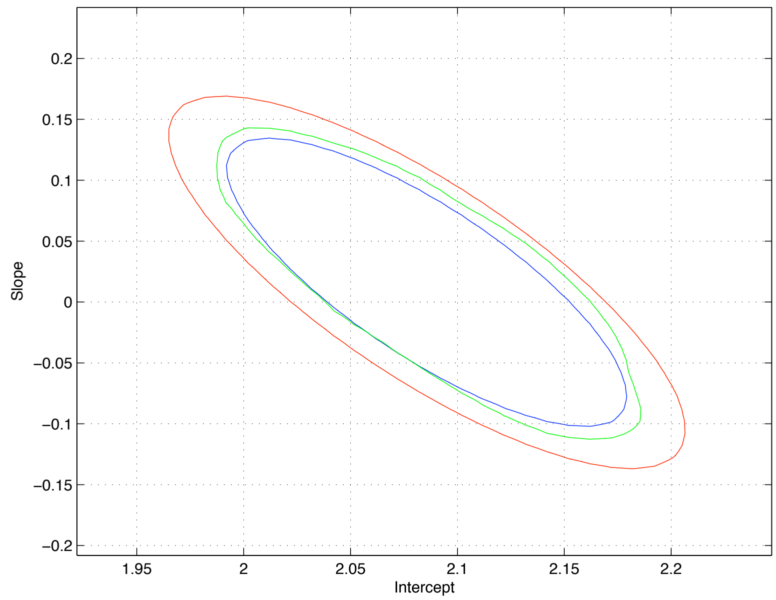

Figure 1: Confidence Regions with No Top-Coding and $\overline{\texttt{wage}} = \$10^8$: Green = Confidence Region for Identifiable Parameters, Red = Confidence Region for Identified Set, Blue = Wald-style Confdience Region

that the results below remain similar for subsample sizes between 20 and 50. Second, when computing critical values, we also used an approximation as described in Remark 2.4 with $B = 200$ because $\binom{n}{b}$ is too large to compute critical values exactly. Finally, following the discussion in Remark 2.7, 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, though the algorithm then requires a few more steps to converge.

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

$$\mathcal{C}_n = S_{11} = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq .0055\} .$$

We also regress Y_i^a on X_i and obtain a Wald-style confidence region. These two confidence regions together with the confidence region for identifiable parameters are displayed in Figure 1. Since the true P generating the data is known, it is also possible to calculate the identified set, which in this case is a singleton. It is given by

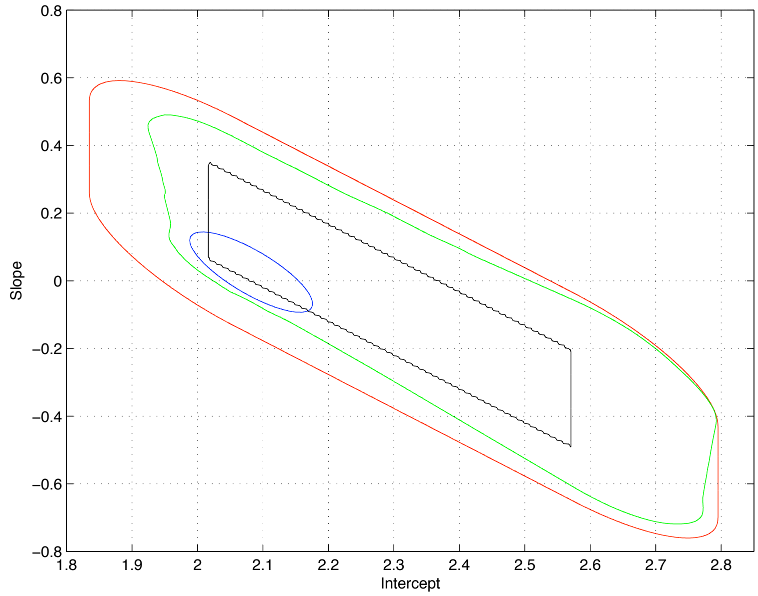

Figure 2: Confidence Regions with 5% Top-Coding and $\overline{\texttt{wage}} = \$10^8$: Green = Confidence Region for Identifiable Parameters, Red = Confidence Region for Identified Set, Blue = Wald-style Confdience Region, Black = Identified Set

$\Theta_0(P) = \{(2.047, .042)\}$. 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 $\underline{\texttt{wage}} = \$41000$, which corresponds to 5% of the population beign subject to top-coding. Algorithm 2.1 converged after 12 steps and the confidence region for the identified set is given by

$$\mathcal{C}_n = S_{12} = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq .0182\} .$$

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

$$\begin{aligned}&\{\theta \in \mathbf{R}^2 : E\{Y_{1,i}|D_i = 0\} \leq \theta_1 \leq E\{Y_{2,i}|D_i = 0\},\\ &E\{Y_{1,i}|D_i = 1\} \leq \theta_1 + \theta_2 \leq E\{Y_{2,i}|D_i = 1\}\}\end{aligned}$$

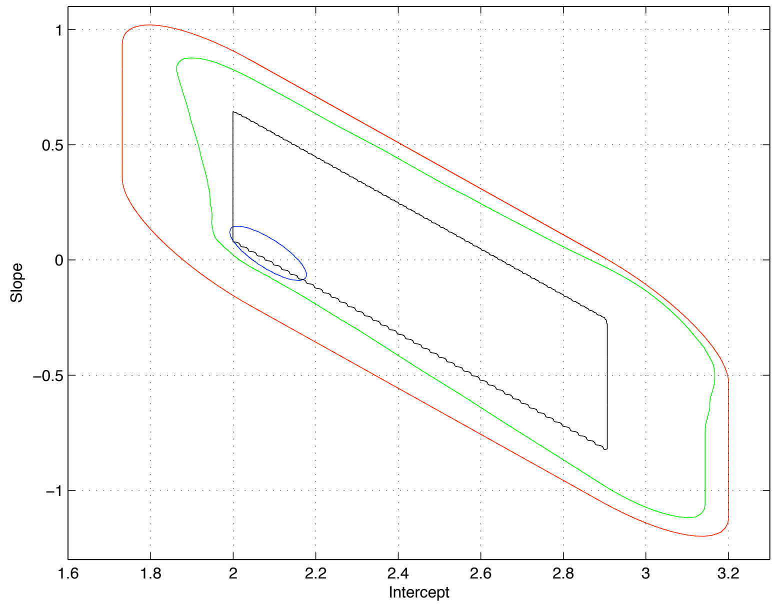

Figure 3: Confidence Regions with 10% Top-Coding and $\overline{\texttt{wage}} = \$10^8$: Green = Confidence Region for Identifiable Parameters, Red = Confidence Region for Identified Set, Blue = Wald-style Confdience Region, Black = Identified Set

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.

In order 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 $\underline{\texttt{wage}}$ to \$35000, which corresponds to 10% of the population begin subject to top-coding. Algorithm 2.1 converged after 9 steps and the confidence region for the identified set is given by

$$\mathcal{C}_n = S_9 = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq .0361\} \ .$$

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

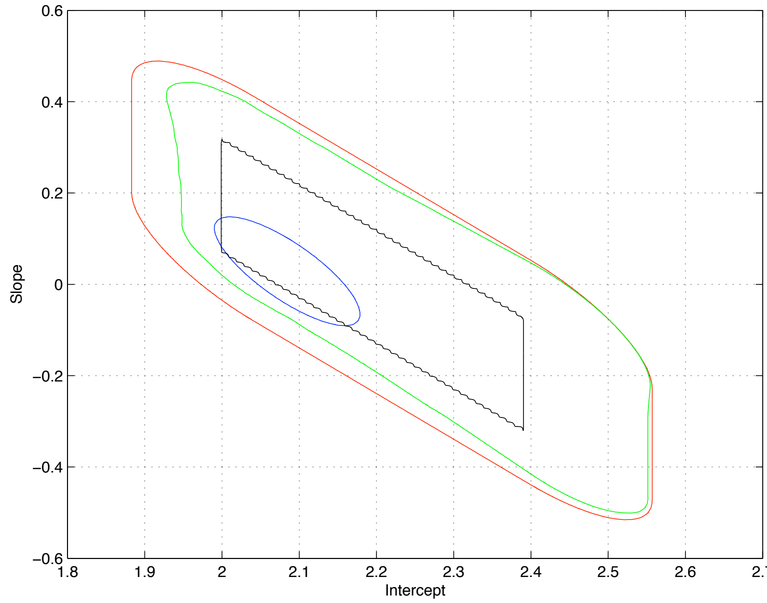

Figure 4: Confidence Regions with 5% Top-Coding and $\overline{\texttt{wage}} = \$10^6$: Green = Confidence Region for Identifiable Parameters, Red = Confidence Region for Identified Set, Blue = Wald-style Confdience Region, Black = Identified Set

confidence region covers an even smaller portion of the identified set, and so inferences based upon it may be even more misleading.

Of course, so far we have assumed a very generous upper bound on annual wages and salaries of $\overline{\texttt{wage}} = \$10^8$. In order to assess how sensitive the qualitative results described above are to the value of $\overline{\texttt{wage}}$, we reexamine the previous case in which 10% of the population is subject to top-coding with the much lower value of $\overline{\texttt{wage}} = \$10^6$. Algorithm 2.1 converged after 12 steps and the confidence region for the identified set is then given by

$$\mathcal{C}_n = S_{12} = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq .0094\} \ .$$

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 $\overline{\texttt{wage}}$ is an important one, as it will impact the sharpness of inferences in such a setting noticeably.

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 studied is the construction of random sets that contain the entire identified set 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 the identified set under a known function. We have verified that these constructions can be applied in several examples of interest.

Our results build upon earlier work by Chernozhukov et al. (2004), who also consider the problem of inference for the same class of partially identified models. An important feature of our procedure for constructing confidence regions for the identified set relative to theirs is that it avoids the need for an initial estimate of the identified set. Moreover, our results provide a justification for iterating their procedure to produce ever smaller confidence regions until a stopping criterion is met while still maintaining the coverage requirement.

For each of our constructions, we have also provided conditions under which our confidence regions are uniformly consistent in level. In the context of confidence regions for identifiable parameters, this issue has been considered earlier by Imbens and Manski (2004), but to the best of our knowledge, we are the first in the literature to consider uniform consistency in level for confidence regions for the identified set.

We have illustrated the use of our methods with an empirical study of the impact of top-coding of outcomes on inferences about the parameters of a linear regression. When confronted with this problem, researchers have typically made strong assumptions about the unobserved values of top-coded outcomes to achieve identification. We have shown that by using the techniques presented in this paper it is possible to weaken these assumptions dramatically and still make meaningful inferences about the parameters of the linear regression. This conclusion, of course, applies more generally to inferences about the parameters of a linear regression when the outcome is interval-censored, which includes top-coded outcomes as a special case.

5 Appendix: Justification of Remark 4.1.2

Theorem 5.1 *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, and suppose $J_n(\cdot, P)$ converges to a limiting distribution $J(\cdot, P)$. 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*

$$L_n(x) = \frac{1}{N_n} \sum_{1 \le i \le N_n} I\{\tau_b(\hat{\vartheta}_{n,b,i} - \vartheta(P)) \le x\} \ , \tag{37}$$

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

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

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

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

Proof: We will only prove (i). The proofs of (ii) and (iii) are similar. Notice that for any $\delta > 0$, $L_n^{-1}(1-\alpha) \ge L_n^{-1}(1-\alpha-\delta)$. Note further that there exists $\delta > 0$ such that $J(\cdot, P)$ is continuous at its $1-\alpha-\delta$ quantile and $J^{-1}(1-\alpha-\delta)$ is arbitrarily close to $J^{-1}(1-\alpha)$. It is worth pointing out that such a δ may not be close to zero. From Theorem 2.1.1 of Politis et al. (1999), we have for such δ that $L_n^{-1}(1-\alpha-\delta) \xrightarrow{P} J^{-1}(1-\alpha-\delta)$. To complete the proof, note that there is $\eta > 0$ arbitrarily small so that $J(\cdot, P)$ is continuous at $J^{-1}(1-\alpha)+\eta$. Thus, by choosing $\eta < \epsilon$ and Slutsky's Theorem, we have that

$$\begin{aligned} & \liminf_{n\to\infty} P\{\tau_n(\hat{\vartheta}_n - \vartheta(P)) \le L_n^{-1}(1-\alpha) + \epsilon\} \\ \ge \ & \liminf_{n\to\infty} P\{\tau_n(\hat{\vartheta}_n - \vartheta(P)) \le J_n^{-1}(1-\alpha) + \eta\} \ge 1-\alpha \ , \end{aligned}$$

as desired. ■

References

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.

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.

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

KATZ, L. F. and AUTOR, D. H. (1999). Changes in the wage structure and earnings inequality. In *Handbook of Labor Economics Volume 3A* (O. Ashenfelter and D. Card, eds.). North Holland, Amsterdam, 1463–1555.

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 identifiable parameters in partially identified econometric models. Technical Report 2006–9, Department of Statistics, Stanford University.

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