Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set

Federico A. Bugni*

Department of Economics
Duke University
federico.bugni@duke.edu

September 11, 2009

ABSTRACT. This paper introduces a novel bootstrap procedure to perform inference in a wide class of partially identified econometric models. We consider econometric models defined by finitely many weak moment inequalities†, which encompass many applications of economic interest. The objective of our inferential procedure is to cover the identified set with a prespecified probability‡.
We compare our bootstrap procedure, a competing asymptotic approximation and subsampling procedures in terms of the rate at which they achieve the desired coverage level, also known as the error in the coverage probability. Under certain conditions, we show that our bootstrap procedure and the asymptotic approximation have the same order of error in the coverage probability, which is smaller than the one obtained by using subsampling. This implies that inference based on our bootstrap and asymptotic approximation should eventually be more precise than inference based on subsampling. A Monte Carlo study confirms this finding in a small sample simulation.

Keywords: Partial Identification, Moment Inequalities, Inference, Bootstrap, Subsampling, Asymptotic Approximation, Rates of Convergence.

*First draft: May, 2006. I am indebted to my advisors, Joel Horowitz, Rosa Matzkin and Elie Tamer for their guidance and support. I thank the co-editor and three anonymous referees for comments and suggestions that have significantly helped to improve this paper. I also thank Donald Andrews, Ivan Canay, Xiaohong Chen, Silvia Glaubach, Nenad Kos, Enno Mammen, Charles Manski, Adam Rosen, Viktor Subbotin, Xun Tang and the participants of seminars at Oberwolfach, Northwestern, UCL, Brown, UPenn, UCLA, Columbia, NYU, Yale, Duke and Michigan for their comments. Financial support from the Robert Eisner Memorial Fellowship and the Dissertation Year Fellowship is gratefully acknowledged. Any and all errors are my own.

†We can also admit models defined by moment equalities by combining pairs of weak moment inequalities.

‡We deal with the objective of covering each element of the identified set with a prespecified probability in Bugni [8].

1 Introduction

This paper contributes to the growing literature on inference in partially identified econometric models. A model is said to be *partially identified* or *set identified* when the sampling process and the maintained assumptions restrict the value of the parameter of interest to a set, called the *identified set*, which is smaller than the logical range of the parameter but potentially[4] larger than a single point. Partially identified models arise naturally in economic models when strong and usually unrealistic assumptions are traded by weaker and more credible ones. In this paper, we consider partially identified models defined by finitely many weak moment inequalities[5].

The goal of this paper is twofold. The first objective is to introduce a novel bootstrap procedure to construct confidence sets for our class of partially identified models. In large samples, our bootstrap procedure achieves exactly the desired coverage probability. The second objective is to compare our bootstrap procedure with competing inferential procedures in terms of the rate of convergence of the error in the coverage probability, that is, in terms of the rate at which they achieve the desired coverage level. To the best of our knowledge, this is the first paper performing this comparison among competing inferential procedures for partially identified models.

The ultimate purpose of our confidence sets is to conduct hypothesis testing in partially identified econometric models. Based on the duality between hypothesis testing and confidence sets, the hypothesis testing problem can be translated into the construction of confidence sets that cover the object of interest with a minimum prespecified probability. In the literature on partially identified models, there are two possible objects of interest, each of them related to different hypothesis testing problems. On the one hand, the object of interest can be the identified set and, on the other hand, the object of interest can be each of the points of the identified set[6]. In this paper, we focus exclusively on the problem of construction of confidence sets that cover the identified set with a prespecified probability[7].

The formal structure of the problem is as follows. An economic model relates the distribution of the observables with a finite dimensional parameter, denoted by θ, that belongs to a parameter space, denoted by Θ. The model is partially identified and hence, the distribution of the observables restricts the true value of the parameter to the identified set, which is denoted by Θ_I. A set $C_n(1-\alpha)$ is a confidence set for the identified set with confidence level $(1-\alpha)$ if and only if the following property is satisfied,

$$\liminf_{n\to+\infty} P\left(\Theta_I \subseteq C_n\left(1-\alpha\right)\right) \geq \left(1-\alpha\right) \tag{1.1}$$

[4]If the parameter of interest is restricted to a single point, the model is said to be point identified.

[5]As mentioned earlier, moment equalities can be admitted by combining pairs of weak moment inequalities.

[6]This distinction was pointed out by Imbens and Manski [17], who show that the confidence set for the identified set will also be a confidence set for each of its elements.

[7]As we mentioned earlier, in Bugni [8], we focus on the problem of construction of confidence sets that cover each of the elements of the identified set with a prespecified probability. In that paper, we propose an analogous bootstrap procedure and obtain the same theoretical results as in this paper.

Romano and Shaikh [24] provide a link between the confidence set $C_n(1-\alpha)$ and a hypothesis testing problem whose null hypothesis is $H_0 : \theta \in \Theta_I$ and whose alternative hypothesis is $H_1 : \theta \notin \Theta_I$. Specifically, if we decide to reject the null hypotheses for all parameter values that lie outside of $C_n(1-\alpha)$, then, in the limit, the probability of incorrectly rejecting at least one of these hypotheses will be less than α.

The confidence set is said to provide *consistent inference in level* if it satisfies the coverage requirement, condition (1.1), with equality. This is a desirable feature since it implies that the confidence set is not excessively large, which would result in unnecessary loss of statistical power of the underlying hypothesis test. We show that our bootstrap procedure provides consistent inference in level.

Our results on confidence regions for partially identified models build upon the criterion function approach introduced by Chernozhukov, Hong and Tamer [10] (henceforth, CHT). In their paper, they implement their inference using a resampling technique called subsampling. In essence, we implement the criterion function approach in a wide class of econometric models using an alternative resampling technique, the bootstrap. Our bootstrap procedure differs qualitatively from replacing the subsampling method provided by CHT [10] with the bootstrap, that is, we do not merely propose a bootstrap analogue of their subsampling method. In fact, we show in section A.2.5 of the appendix that a bootstrap analogue of their subsampling procedure would, in general, fail to be consistent in level. The difference between our bootstrap method and the bootstrap analogue of the subsampling procedure proposed by CHT [10] lies in the choice of the bootstrap criterion function, which is the key to our consistency result. Following similar techniques to those used to implement our bootstrap scheme, we also propose an asymptotic approximation to perform consistent inference in level[8].

There are currently many methods available to implement inference in partially identified econometric models. Given the choice of the criterion function, the researcher can implement inference using our bootstrap, asymptotic approximation or subsampling. Since all these methods provide consistent inference in level (that is, they all achieve the desired goal, asymptotically), an important basis of comparison is the rate at which the error in the coverage probability vanishes (that is, the rate at which the goal is achieved). If two methods have errors in the coverage probability that converge to zero at different rates, then the one that converges faster will eventually be more accurate than the one that converges slower. One of the main contributions of this paper is to show that, under certain conditions, our bootstrap and our asymptotic approximation have errors in the coverage probability that converge to zero at the same rate, which is a faster rate than the one obtained by using subsampling. Hence, under these conditions, our results imply that inference based on our bootstrap and our asymptotic approximation should eventually be more precise than inference based on subsampling[9].

---

[8] In independent research, a similar asymptotic approximation has also been proposed by Soares [28], CHT [10] and Andrews and Soares [3].

[9] This result can be related to results in the literature on inference about sample averages, where the rate of con-

Several papers have proposed different inferential methods to perform inference in partially identified models. Most of these references only construct confidence sets that cover the elements of the identified set with a prespecified probability, which is a different coverage objective from the one considered in this paper[10]. Notable exceptions are CHT [10], Beresteanu and Molinari [5] and Romano and Shaikh [24]. Beresteanu and Molinari [5] propose an alternative approach to inference in partially identified models using random set theory, which is significantly different from the techniques used in our paper. Romano and Shaikh [24] consider the general problem of constructing coverage regions for the identified set using a subsampling stepdown control procedure that is comparable to the subsampling procedure proposed by CHT [10]. They show formally that the subsampling stepdown control procedure cannot be replaced by a naive bootstrap stepdown control procedure. All of the procedures considered in our paper, including our bootstrap approximation, can be used as an ingredient in a stepdown control inferential procedure that results in consistent inference in level[11].

The rest of the paper is organized as follows. Section 2 considers the construction of confidence sets for the identified set. Section 2.1 introduces our assumptions and provides an example of an econometric model where these assumptions are satisfied. In section 2.2, we introduce our bootstrap procedure to perform inference, demonstrate its consistency in level and analyze its error in the coverage probability. In section 2.3, we consider the two competing inferential procedures, subsampling and asymptotic approximation, for which we also show consistency in level and analyze the error in coverage probability. Section 3 concludes the paper and provides directions for further research. The online appendix collects all the proofs of the paper, intermediate results, supplementary explanations and Monte Carlo simulations.

2 Confidence sets for the identified set

Our objective is to construct confidence sets that cover the identified set with a prespecified probability and we choose to accomplish this objective using the criterion function approach. According to this approach, the first step is to define a non-negative function of the parameter space, denoted by Q, that equals zero if and only if the parameter belongs to the identified set. This function is referred to as *criterion function* because it provides a criterion that characterizes the identified set. We denote its sample analogue by Q_n and define $\Gamma_n = \sup_{\theta \in \Theta_I} a_n Q_n(\theta)$, where $\{a_n\}_{n=1}^{+\infty}$ is a sequence of constants that makes the (asymptotic) distribution of Γ_n non-degenerate. If we let c_n denote the $(1-\alpha)$ quantile of Γ_n, then, a confidence set of the identified set with

vergence of subsampling procedures is likely to be slow, relative to the bootstrap or the asymptotic approximation. See, for example, Horowitz [16], Politis and Romano [22] and Politis, Romano and Wolf [23].

[10] These include Rosen [26], Pakes, Porter, Ho and Ishii [21], Andrews, Berry and Jia [2], Canay [9], Romano and Shaikh [25], Soares [28] and Andrews and Soares [3].

[11] I thank an anonymous referee for suggesting this point. This is formally shown in section A.4.1 of the appendix.

confidence level $(1-\alpha)$ is given by,

$$C_n(1-\alpha) = \{\theta \in \Theta : a_n Q_n(\theta) \leq c_n\} \tag{2.2}$$

Therefore, the criterion function approach translates the problem of constructing confidence sets into a problem of approximation of the quantiles of Γ_n. This approximation problem is non-standard precisely because the econometric model is partially identified. In a class of models defined by moment inequalities and equalities, we show how to perform this approximation using the bootstrap.

2.1 Setup

In this section, we introduce the assumptions that define our econometric model. We consider two separate sets of assumptions. The first set of assumptions will constitute what we call *the general model*. The second set of assumptions will be a particular subset of the first one and will constitute what we call *the conditionally separable model*. The reason to consider two separate setups is that consistency in level can be shown under the assumptions of the general model but results regarding rates of convergence require the stronger restrictions imposed by the conditionally separable model. After introducing both sets of assumptions, we provide an example to illustrate each of these frameworks.

2.1.1 General model

The following assumptions constitute our general model in the independent and identically distributed (i.i.d.) setting.

(A1) For the probability space $(\Omega, \mathcal{B}, \mathbf{P})$, let $Z : \Omega \to S_Z$ be a random vector. We observe an i.i.d. sample of size n, denoted by $\mathcal{X}_n = \{Z_i\}_{i=1}^n$.

(A2) The parameter space, denoted by Θ, is a compact and convex subset of a finite dimensional Euclidean space $\mathbb{R}^\eta$ $(\eta < +\infty)$.

(A3) The identified set, denoted by Θ_I, is given by,

$$\Theta_I = \left\{\theta \in \Theta : \{\mathbb{E}(m_j(Z,\theta)) \leq 0\}_{j=1}^J\right\}$$

where $m(z,\theta) : S_Z \times \Theta \to \mathbb{R}^J$ is a (jointly) measurable function and $\mathbb{E}(m(Z,\theta)) : \Theta \to \mathbb{R}^J$ is a lower semi-continuous function. Moreover, Θ_I is a proper subset of Θ.

(A4) For every $\theta \in \Theta$ and every $j = 1, 2, ..., J$, the variance of $m_j(Z,\theta)$ is finite and positive. For every $z \in S_Z$, the function $\left\{m(z,\theta) - \mathbb{E}(m(Z,\theta)) : \Theta \to \mathbb{R}^J\right\}$ belongs to B, which is a separable subset of the space of bounded functions that map Θ onto $\mathbb{R}^J$ with the sup-norm metric. The empirical process $v_n(m_\theta) = n^{-1/2}\sum_{i=1}^{n}(m(Z_i,\theta) - \mathbb{E}(m(Z,\theta)))$ is stochastically equicontinuous, that is, for any $\varepsilon > 0$,

$$\lim_{\eta\downarrow 0}\limsup_{n\to+\infty} P^*\left(\sup_{\theta\in\Theta}\sup_{\{\theta'\in\Theta:\|\theta'-\theta\|\leq\eta\}}\|v_n(m_\theta) - v_n(m_{\theta'})\| > \varepsilon\right) = 0$$

where $\|\cdot\|$ denotes Euclidean distance and P^* denotes the outer measure[12] with respect to P.

We briefly comment on some of the assumptions. Assumption (A1) requires that the sample is i.i.d.. The result of consistency of the bootstrap procedure proposed in this paper is based on the law of large numbers, the central limit theorem and the law of iterated logarithm applied to empirical processes. Consistency of our bootstrap procedure can be generalized to non i.i.d. settings, provided that these results hold and, of course, that the resampling method is adequately modified for these settings.

Assumption (A3) defines the identified set as the intersection of finitely many weak moment inequalities. Of course, equality restrictions can be accommodated by combining two moment inequalities. Also, notice that assumption (A3) allows the identified set to be empty, which would imply that the model is misspecified. The present setup allows for econometric models defined by *conditional* moment conditions as long as the conditioning covariates have finite support[13]. To see why, suppose that the conditioning covariate X has finite support given by S_X and let the identified set be given by, $\Theta_I = \{\theta\in\Theta : \{\{\mathbb{E}(M_j(Y,\theta)|X=x)\leq 0\}_{j=1}^{J}\}_{x\in S_X}\}$, where $M(y,\theta) : S_Y\times\Theta\to\mathbb{R}^J$ is a jointly measurable function and, for every $x\in S_X$, $\mathbb{E}(M(Y,\theta)|X=x) : \Theta\to\mathbb{R}^J$ is lower semi-continuous. By defining $Z = (Y,X)$ and $m(Z,\theta) = M(Y,\theta)1[X=x]$, the identified set is re-expressed according to assumption (A3).

When we combine the total boundedness of the parameter space of assumption (A2) with the stochastic equicontinuity condition of assumption (A4), we deduce that the class of functions $\{m(z,\theta) - \mathbb{E}(m(Z,\theta)) : S_Z\to\mathbb{R}^J\}$, indexed by $\theta\in\Theta$, is P-Donsker.

2.1.2 Conditionally separable model

The following assumptions constitute our conditionally separable model in the i.i.d. setting.

(B1) For the probability space $(\Omega,\mathcal{B},\mathbf{P})$, let $(X,Y) : \Omega\to S_X\times\mathbb{R}^J$ be a random vector, where the support of X, denoted S_X, is composed of K values: $S_X = \{x_k\}_{k=1}^{K}$. We observe an i.i.d. sample of size n, denoted by $\mathcal{X}_n = \{X_i,Y_i\}_{i=1}^{n}$.

[12] Let $(\Omega,\mathcal{B},\mathbf{P})$ be a probability space. For any arbitrary subset of Ω, denoted A, its outer measure is defined by $P^*(A) = \inf_{S\subseteq\mathcal{B}}\{P(S) : A\subseteq S\}$.

[13] The methods proposed in this paper cannot handle conditioning covariates with infinite support. If this is the case, one can still use our techniques by partitioning the support of the conditioning covariate into finitely many bins. In this process, some information will be lost, and so our method will result in conservative inference.

(B2) The parameter space, denoted by Θ, is a compact and convex subset of a finite dimensional Euclidean space $\mathbb{R}^{\eta}$ ($\eta < +\infty$).

(B3) The identified set, denoted by Θ_I, is given by,

$$\Theta_I = \left\{ \theta \in \Theta : \left\{ \left\{ \mathbb{E}\left(Y_j - M_j\left(\theta, X\right) | X = x_k \right) \leq 0 \right\}_{j=1}^{J} \right\}_{k=1}^{K} \right\}$$

where, for each $x \in S_X$, $M\left(\theta, x\right) : \Theta \rightarrow \mathbb{R}^J$ is continuous. Moreover, Θ_I is a proper subset of Θ.

(B4) For every $x \in S_X$ and for every $j = 1, 2, ..., J$, $\{Y_j | X = x\}$ has positive variance and finite fourth absolute moments.

This setup is a particular case of the general model that strengthens assumptions (A1), (A3) and (A4). Assumption (B1) organizes the observable variables into conditioning covariates and dependent variables, and requires the former to have finite support.

Recall that assumption (A3) defined the identified set as the intersection of moment inequalities of the form $\mathbb{E}\left(m_j\left(Z, \theta\right) | X = x\right) \leq 0$. Assumption (B3) strengthens this by requiring that the conditional expectation $\mathbb{E}\left(m_j\left(Z, \theta\right) | X = x\right)$ can be separated into the expectation of a random variable that does not involve the parameter, $\mathbb{E}\left(Y_j | X = x\right)$, and a conditionally non-stochastic function of the parameter, $\mathbb{E}\left(M_j\left(\theta, X\right) | X = x\right)$. Finally, lemma A.1 of the appendix shows that assumptions (B1)-(B4) imply assumption (A4).

2.1.3 Illustrative Example

In order to illustrate both of our frameworks and highlight their differences, we consider the following example related to Manski [20]. Suppose that our model predicts that $\mathbb{E}\left(Z - f\left(X, \theta\right) | W\right) = 0$, where f is a known function, Z is the explained variable, X is the explanatory variable, θ is the parameter of interest and W is an exogenous variable. Typical examples of this setup are linear index models, such as the linear model, the probit model or the logit model.

Suppose that certain observations of the explained variable are missing (or censored). Let U denote the binary variable that takes value one if the observation is unobserved and zero otherwise. Suppose that $\{Z|W\}$ has logical lower and upper bounds, given by $Z_L\left(W\right)$ and $Z_H\left(W\right)$, respectively[14]. Also assume that the support of W is given by finitely many values: $S_W = \{w_k\}_{k=1}^{K}$. Under these conditions, the identified set for the parameter of interest is given by,

$$\Theta_I = \left\{ \theta \in \Theta : \left\{ \begin{array}{c} \mathbb{E}\left(-\left(Z\left(1-U\right) + Z_H\left(w_k\right)U - f\left(X, \theta\right)\right) 1\left[W = w_k\right]\right) \leq 0 \\ \mathbb{E}\left(\left(Z\left(1-U\right) + Z_L\left(w_k\right)U - f\left(X, \theta\right)\right) 1\left[W = w_k\right]\right) \leq 0 \end{array} \right\}_{k=1}^{K} \right\}$$

[14] When the event has no logical lower bound (respectively, upper bound), then $Z_L\left(W\right) = -\infty$ (respectively, $Z_H\left(W\right) = +\infty$).

Under random sampling and certain regularity conditions[15], this example satisfies the assumptions of the general framework. In particular, if the explanatory variable X is endogenous or has continuous support, it cannot replace W in role of conditioning covariate and therefore, the example does not satisfy the assumptions of the conditionally separable model. On the other hand, if the explanatory variable X is exogenous and has finite support, it can replace W as the conditioning covariate and the example satisfies the assumptions of the conditionally separable model.

2.2 Bootstrap procedure

In this section, we introduce our bootstrap procedure to construct confidence regions for the identified set. In order to implement any inferential procedure based on the criterion function approach, we need to complete certain steps. First, we need to define the criterion function for our problem. Second, we need to generate an estimator of the identified set. This estimator is not our final goal, but an input to our inferential procedure. Once we completed these preliminary steps, we are ready to define the resampling procedure that implements our inference.

2.2.1 Criterion function

By definition, a function $Q : \Theta \to \mathbb{R}$ is a valid criterion function if it is non-negative and takes value zero if and only if it is evaluated at a parameter in the identified set. Lemma A.2 of the appendix characterizes all possible criterion functions for the type of models considered in this paper. This result reveals that there is a wide range of possible criterion functions. By restricting the class of functions under consideration, we can obtain desirable asymptotic results for our inferential procedures, such as consistency in level and rates of convergence. With this objective in mind, we consider the following assumption.

Assumption (CF) The population criterion function is given by one of the following functions: $Q(\theta) = \sum_{j=1}^{J} w_j [\mathbb{E}(m_j(Z,\theta))]_+$ or $Q(\theta) = \max\{w_j[\mathbb{E}(m_j(Z,\theta))]_+\}_{j=1}^{J}$, where $\{w_j\}_{j=1}^{J}$ are (arbitrary) positive constants.

Throughout this paper, we will focus on criterion functions that satisfy assumption (CF), but most of our results extend to more general criterion functions[16].

2.2.2 Estimation of the identified set

The estimator of the identified set is an ingredient in the construction of confidence sets that cover the identified set with a prespecified probability. An estimator of the identified set is adequate

[15] These conditions are made explicit in section A.2.2.

[16] In particular, our theoretical results extend to a generalization of assumption (CF), called assumption (CF'), which is described in section A.2.3 of the appendix.

for the purpose of inference if a confidence set which uses this estimator as an input produces inference that is consistent in level.

By definition, the identified set is the subset of the parameter space that satisfies $Q(\theta) = 0$. Therefore, the *analogy principle* suggests defining the estimator of the identified set as the collection of parameters that satisfy $Q_n(\theta) = 0$. In our context, this set estimator would be given by, $\hat{\Theta}_I^{AP} = \{\theta \in \Theta : \{\mathbb{E}_n(m_j(Z,\theta)) \leq 0\}_{j=1}^J\}$, where for every $j = 1, 2, ..., J$, $\mathbb{E}_n(m_j(Z,\theta))$ denotes $n^{-1}\sum_{i=1}^n m_j(Z_i,\theta)$. This random set is called *the analogy principle estimator.*

In settings of practical relevance, the analogy principle estimator is not an adequate estimator for the purpose of inference. Following CHT [10], an adequate estimator requires the introduction of some slackness in the sample moment inequalities, thereby expanding the analogy principle estimator. Specifically, if we let $\{\tau_n\}_{n=1}^{+\infty}$ be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln \ln n}/\tau_n = o(1)$ (almost surely), then our estimator of the identified set is given by,

$$\hat{\Theta}_I(\tau_n) = \left\{\theta \in \Theta : \left\{\mathbb{E}_n(m_j(Z,\theta)) \leq \tau_n/\sqrt{n}\right\}_{j=1}^J\right\}$$

Our estimator can be interpreted as an expansion of the analogy principle estimator by a slackness factor of $\tau_n/\sqrt{n}$. The following lemma formalizes its the properties.

Lemma 2.1 *Assume (A1)-(A4). Let $\{\tau_n\}_{n=1}^{+\infty}$ be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln \ln n}/\tau_n = o(1)$, almost surely, and define $\hat{\Theta}_I(\tau_n) = \{\theta \in \Theta : \{\mathbb{E}_n(m_j(Z,\theta)) \leq \tau_n/\sqrt{n}\}_{j=1}^J\}$. For a sequence of positive numbers $\{\varepsilon_n\}_{n=1}^{+\infty}$ such that $\varepsilon_n = o(1)$ and $(\tau_n/\sqrt{n})/\varepsilon_n = o(1)$, almost surely, define $\Theta_I(\varepsilon_n) = \{\theta \in \Theta : \{\mathbb{E}(m_j(Z,\theta)) \leq \varepsilon_n\}_{j=1}^J\}$.*

If the identified set is non-empty then, $\liminf\{\Theta_I \subseteq \hat{\Theta}_I(\tau_n) \subseteq \Theta_I(\varepsilon_n)\}$, *almost surely, and if the identified set is empty then,* $\liminf\{\hat{\Theta}_I(\tau_n) = \varnothing\}$, *almost surely.*

When the identified set is non-empty, our set estimate will eventually be "sandwiched" between two sets, almost surely. These sets are the identified set and a set that converges to the identified set. When the identified set is empty, our set estimate will eventually become empty, almost surely. As we show later, lemma 2.1 implies that the proposed estimator of the identified set is adequate for the purpose of inference.

The restrictions on the sequence $\{\tau_n\}_{n=1}^{+\infty}$ provide little guidance on how to implement the estimator (and the inference based on it) in a finite sample setting. We comment on this important practical question in the next subsection.

2.2.3 The procedure

We now introduce our bootstrap procedure to construct confidence sets that cover the identified set with a prespecified probability. We will propose two different procedures: one to be used if the model satisfies the assumptions of the general model and one to be used exclusively if the model satisfies the assumptions of the conditionally separable model.

Bootstrap procedure for the general model The following bootstrap method is intended for the general model, and so, in particular, it could also be applied to the conditionally separable model[17]. This will be referred to as the bootstrap procedure *for the general model* and it consists of the following steps:

1. Choose $\{\tau_n\}_{n=1}^{+\infty}$ to be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln\ln n}/\tau_n = o(1)$, almost surely,

2. Estimate the identified set with $\hat{\Theta}_I(\tau_n) = \left\{\theta \in \Theta : \{\mathbb{E}_n(m_j(Z,\theta)) \leq \tau_n/\sqrt{n}\}_{j=1}^{J}\right\}$,

3. Repeat the following for $s = 1, 2, ..., S$. Construct bootstrap samples of size n, by sampling randomly with replacement from the data. Denote the bootstrapped observations by $\{Z_i^*\}_{i=1}^n$ and, for every $j = 1, 2, ..., J$, let $\mathbb{E}_n^*(m_j(Z,\theta))$ denote $n^{-1}\sum_{i=1}^n m_j(Z_i^*,\theta)$. Compute,

$$\Gamma_n^* = \begin{cases} \sup_{\theta \in \hat{\Theta}_I(\tau_n)} G\left(\left\{\begin{array}{c} [\sqrt{n}(\mathbb{E}_n^*(m_j(Z,\theta)) - \mathbb{E}_n(m_j(Z,\theta)))]_+ \\ *1[|\mathbb{E}_n(m_j(Z,\theta))| \leq \tau_n/\sqrt{n}] \end{array}\right\}_{j=1}^{J}\right) & \text{if } \hat{\Theta}_I(\tau_n) \neq \varnothing \\ 0 & \text{if } \hat{\Theta}_I(\tau_n) = \varnothing \end{cases}$$

4. Let $\hat{c}_n^B(1-\alpha)$ be the $(1-\alpha)$ quantile of the distribution of Γ_n^*, approximated with arbitrary accuracy in the previous step. The $(1-\alpha)$ confidence set for the identified set is given by $\hat{C}_n^B(1-\alpha) = \left\{\theta \in \Theta : \sqrt{n}Q_n(\theta) \leq \hat{c}_n^B(1-\alpha)\right\}$.

In order to implement our procedure, we need to specify the sequence $\{\tau_n\}_{n=1}^{+\infty}$ described in the first step. This sequence enters the procedure in two places. First, it enters in the estimation of the identified set in step 2 and, second, it enters in the indicator function term in the bootstrap criterion function in step 3[18]. The restrictions on the rate of the sequence $\{\tau_n\}_{n=1}^{+\infty}$ in step 1 provide little guidance on how to choose this sequence in practice. As we will show later, our bootstrap procedure will be consistent in level and have the same rate of convergence regardless of the specific choice of the sequence $\{\tau_n\}_{n=1}^{+\infty}$. Thus, our asymptotic analysis does not provide a criterion for an "optimal" choice of this sequence. In this sense, we consider that the development of a data-dependent choice of this sequence is out of the scope of this paper. The experience drawn from Monte Carlo simulations seems to indicate that the finite sample performance of our inferential method does not depend critically on the specific choice of this sequence.

The key to the consistency in level of our bootstrap procedure is the bootstrap analogue criterion function defined in step 3. In particular, it is essential to the consistency result that we introduce (a) the recentering term, that is, subtracting the sample moment from the bootstrap

[17]Nevertheless, as we will explain next, if the model satisfies the assumptions of the conditionally separable model there are advantages of using the bootstrap procedure specialized for this framework.

[18]In principle, the sequence $\{\tau_n\}_{n=1}^{+\infty}$ in these two steps could be two different sequences provided that they both satisfy the rate requirements of step 1. The formal arguments in the appendix allow these two sequences to differ. We restrict both sequences to coincide in the main text only to simplify the notation.

sample moment and (b) the indicator function term. Because of these two terms, our bootstrap procedure differs qualitatively from the bootstrap version of the subsampling scheme proposed by CHT [10]. We analyze these differences in section A.2.5 of the appendix.

Bootstrap procedure for the conditionally separable model In the conditionally separable model, we will not show the consistency in level of the inferential procedure, but we will also establish the rate of convergence of the error of the bootstrap approximation.

In order to understand why we need to introduce a separate bootstrap method for the conditionally separable model, we need to distinguish between whether the design of the covariates is fixed or random. The design of the covariates refers to how the econometrician perceives the distribution of covariates in the sample. If the design of the covariates is fixed, then the distribution of the covariates is considered to be non-stochastic (or stochastic and conditioned upon) and if the design of the covariates is random, then the distribution of the covariates is considered to be stochastic. Of course, the inferential procedure we perform is different depending on the case.

In the fixed covariates case, the cell frequency of the covariates is constant and this implies that the bootstrap procedure proposed in the previous section delivers a rate of convergence of order $n^{-1/2}$. In the random covariates case, the cell frequency of the covariates is random and this implies that the bootstrap procedure proposed in the previous section delivers a rate of convergence of order $n^{-1/2}\ln n \ln\ln n$ (rather than $n^{-1/2}$)[19]. Nevertheless, it is possible to design a bootstrap method which can achieve a rate of convergence of order $n^{-1/2}$ independently of the design of the covariates. This is referred to as *the bootstrap procedure specialized for the conditionally separable model* and it consists of the following steps:

1. Choose $\{\tau_n\}_{n=1}^{+\infty}$ to be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln\ln n}/\tau_n = o(1)$, almost surely,

2. Estimate the identified set with $\hat{\Theta}_I(\tau_n) = \left\{\theta \in \Theta : \left\{\left\{\hat{p}_k\left(\mathbb{E}_n(Y_j|x_k) - M_j(\theta, x_k)\right) \leq \tau_n/\sqrt{n}\right\}_{j=1}^{J}\right\}_{k=1}^{K}\right\}$,

3. Repeat the following for $s = 1, 2, ..., S$. Construct bootstrap samples of size n, by sampling randomly with replacement from the data[20]. Denote the bootstrapped observations by $\{Y_i^*, X_i^*\}_{i=1}^n$, and, for every $k = 1, 2, ..., K$ and $j = 1, 2, ..., J$, let $\hat{p}_k^* = n^{-1}\sum_{i=1}^n 1[X_i^* = x_k]$ and $\mathbb{E}_n^*(Y_j|x_k) = (\hat{p}_k^* n)^{-1}\sum_{i=1}^n Y_{j,i}^* 1[X_i^* = x_k]$. Compute,

$$\Gamma_n^* = \begin{cases} \sup_{\theta \in \hat{\Theta}_I(\tau_n)} G\left(\left\{\left\{\begin{array}{c} [\sqrt{n}\hat{p}_k^*\left(\mathbb{E}_n^*(Y_j|x_k) - \mathbb{E}_n(Y_j|x_k)\right)]_+ * \\ 1\left[\left|\hat{p}_k\left(\mathbb{E}_n(Y_j|x_k) - M_j(\theta, x_k)\right)\right| \leq \tau_n/\sqrt{n}\right] \end{array}\right\}_{j=1}^{J}\right\}_{k=1}^{K}\right) & \text{if } \hat{\Theta}_I(\tau_n) \neq \varnothing \\ 0 & \text{if } \hat{\Theta}_I(\tau_n) = \varnothing \end{cases}$$

[19] This result is available from the author, upon request.

[20] Of course, bootstrap samples are constructed to respect the assumption on the design of the covariates.

4. Let $\hat{c}_n^B(1-\alpha)$ be the $(1-\alpha)$ quantile of the distribution of Γ_n^*, simulated with arbitrary accuracy in the previous step. The $(1-\alpha)$ confidence set for the identified set is given by $\hat{C}_n^B(1-\alpha) = \left\{\theta \in \Theta : \sqrt{n}Q_n(\theta) \leq \hat{c}_n^B(1-\alpha)\right\}$.

The only distinction between the general procedure and the one specialized for the conditionally separable model occurs in the definition of the bootstrap criterion function in step 3. In the latter one, the argument inside the $[\cdot]_+$ function is a random variable rather than a random function, which is the key feature that allows us to obtain a rate of convergence of order $n^{-1/2}$, regardless of the assumption about the covariates.

2.2.4 Asymptotic properties

In this section, we establish the asymptotic properties of our bootstrap procedure. The results of this section are based on two representation theorems, which are stated and proved in the appendix. As a first step, we show that the distribution of the statistic of interest has a certain asymptotic representation (theorem A.1). In a second step, we establish that, conditional on the sample, our bootstrap approximation has an analogous asymptotic representation (theorem A.3). Based on these results, we can show that bootstrap confidence sets are consistent in level.

Theorem 2.1 *[Consistency in level - bootstrap approximation] Assume (A1)-(A4) and (CF). If the identified set is non-empty then, for any* $\alpha \in (0, 0.5)$,

$$\lim_{n \to +\infty} P\left(\Theta_I \subseteq \hat{C}_n^B(1-\alpha)\right) = (1-\alpha)$$

The representation theorems are also the key to analyze the *error in the coverage probability* (ECP) of the bootstrap approximation, which is the difference between the desired coverage level and the actual coverage level. By theorem 2.1, the error in the coverage probability of our bootstrap approximation converges to zero. In the conditionally separable model, the representation theorems are used to provide an upper bound on the rate at which this convergence occurs.

Theorem 2.2 *[ECP - bootstrap approximation] Assume (B1)-(B4), (CF) and choose the bootstrap procedure to be the one specialized for the conditionally separable model. If the identified set is non-empty then, for any* $\alpha \in (0, 0.5)$,

$$\left|P\left(\Theta_I \subseteq \hat{C}_n^B(1-\alpha)\right) - (1-\alpha)\right| = O\left(n^{-1/2}\right)$$

In terms of coverage, the only relevant case is the non-empty identified set, since the empty set is trivially covered by any confidence set[21]. The previous theorem shows that, in the conditionally separable model, the error in the coverage probability converges to zero at a rate of order $n^{-1/2}$.

[21] Our bootstrap procedure also exhibits desirable properties when the identified set is empty. This is established in lemma A.7 in the appendix.

Provided that the sequence $\{\tau_n\}_{n=1}^{+\infty}$ satisfies the requirements of step 1, the rate of convergence of the error in the coverage probability does not depend on the particular choice of the rate. In this sense, our asymptotic analysis does not provide a criterion for an "optimal" choice of this sequence.

2.3 Alternative procedures

In the previous sections, we proposed a bootstrap scheme to perform inference in partially identified models and we studied its properties. In this section, we consider alternative inferential methods.

2.3.1 Subsampling

There are different subsampling procedures that can be proposed to approximate the distribution of interest. A first scheme would be a subsampling analogue of the bootstrap procedure proposed in the preceding section, which will be referred to as subsampling 1. A second scheme[22] would be the subsampling procedure proposed by CHT [10] which is referred to as subsampling 2. Both of these procedures are described and studied in detail in section A.6.1 of the appendix. The basic difference between the two methods is the definition of the subsampling criterion function in step 3. On the one hand, subsampling 1 proposes a criterion function that includes a recentering term and an indicator function term, exactly as in the bootstrap scheme of section 2.2.3. On the other hand, the criterion function of subsampling 2 has none of these terms. The benchmark subsampling scheme for the presentation of the results is subsampling 1, but, at the end of this subsection, we will briefly discuss how the results differ from those of subsampling 2.

First, we establish the consistency in level of the benchmark subsampling approximation.

Theorem 2.3 *[Consistency in level - subsampling approximation 1] Assume (A1)-(A4) and (CF). If the identified set is non-empty then, for any $\alpha \in (0, 0.5)$,*

$$\lim_{n \to +\infty} P\left(\Theta_I \subseteq \hat{C}^{SS_1}_{b_n,n}(1-\alpha)\right) = (1-\alpha)$$

Second, we establish the rate of convergence of the error in the coverage probability of the inference based on the benchmark subsampling approximation.

Theorem 2.4 *[ECP - subsampling approximation 1] Assume (B1)-(B4), (CF) and that the distribution of $\{\{Y_j 1[X = x_k]\}_{j=1}^J\}_{k=1}^K$ is strongly non-lattice. If the identified set is non-empty then, for any $\alpha \in (0, 0.5)$,*

$$\left| P\left(\Theta_I \subseteq \hat{C}^{SS_1}_{b_n,n}(1-\alpha)\right) - (1-\alpha) \right| = O\left(b_n/n + b_n^{-1/2}\right)$$

[22] I thank an anonymous referee for suggesting the consideration of the second version of subsampling.

This theorem establishes an upper bound on the rate at which the error in coverage probability of the subsampling approximation converges to zero. This upper bound depends on the choice of the subsampling size, reflecting the usual trade-off: increasing subsampling size increases the precision of the averages within a subsample but decreases the total number of subsamples available. The choice of the subsampling size that minimizes this upper bound is $b_n = O\left(n^{2/3}\right)$, which results in error in the coverage probability of order $n^{-1/3}$.

We can provide conditions under which this rate constitutes not just an upper bound on the rate of convergence of the error in the coverage probability, but also a lower bound. We now describe the results but all the formal arguments are provided in section A.6.1 of the appendix. Under certain conditions, lemma A.9 in the appendix shows that the conditional distribution of our subsampling approximation has the following asymptotic representation,

$$\underbrace{P\left(\hat{\Gamma}^{SS_1}_{b_n,n} \leq h|\mathcal{X}_n\right)}_{\text{Subsampling approx.}} = \underbrace{P\left(\Gamma_n \leq h\right)}_{\text{Exact distribution}} + K_1\left(h\right) b_n^{-1/2} + K_2\left(h\right)\frac{b_n}{n} + o_p\left(\frac{b_n}{n} + b_n^{-1/2}\right) \tag{2.3}$$

uniformly over $h \geq \mu$ (for any $\mu > 0$), where K_1 and K_2 are two non-stochastic functions given in the appendix. Politis, Romano and Wolf [23] provide intuition for the nature of the two leading terms in the error of the subsampling approximation. The term of order $b_n^{-1/2}$ appears because we are approximating a distribution by repeatedly extracting random samples of size b_n from the data. This leading term will also appear in the bootstrap, but will be order $n^{-1/2}$ instead of $b_n^{-1/2}$. The term of order b_n/n appears because samples are extracted without replacement from a finite population, which introduces error in the approximation of the variance of the distribution. This term does not appear in the bootstrap, where samples are extracted with replacement.

From this equation, if follows that, for any $h \geq \mu$, the absolute value of the error in the subsampling approximation 1 is minimized by choosing subsampling size $b_n = C\left(h\right) n^{2/3}$, where $C\left(h\right)$ is the positive minimizer of $|K_1\left(h\right) C\left(h\right)^{-1/2} + K_2\left(h\right) C\left(h\right)|$. Therefore, if the functions K_1 and K_2 share the sign (for relevant values of h), then the error in the subsampling approximation 1 converges to zero at a rate that cannot be faster than $n^{-1/3}$.

For the purpose of inference, we are interested in values of h in a neighborhood of the $(1-\alpha)$ quantile of the limiting distribution, which we denote by $c_\infty\left(1-\alpha\right)$. In lemma A.11 in the appendix, we show that for all significance levels that are relevant for inferential purposes, $K_2\left(c_\infty\left(1-\alpha\right)\right)$ is positive. Therefore, if $K_1\left(c_\infty\left(1-\alpha\right)\right)$ is also positive, the previous logic implies that the subsampling approximation 1 converges to the distribution of interest at an exact rate of $n^{-1/3}$. This is the content of theorem A.11 in the appendix. The conditions under which $K_1\left(c_\infty\left(1-\alpha\right)\right)$ is positive involve restrictions on the moments of $\{\{Y_j 1\left[X = x_k\right]\}_{j=1}^{J}\}_{k=1}^{K}$ that, to the best of our knowledge, lack intuitive interpretation. In the case that $K_1\left(c_\infty\left(1-\alpha\right)\right)$ is non-positive, it might be possible to set the right hand side of equation (2.3) to be $o_p\left(n^{-1/3}\right)$ by a particularly judicious choice of $C\left(h\right)$. However, this approach would not be very practical, since it requires careful empirical selection of the subsampling size based on estimation of $K_1\left(c_\infty\left(1-\alpha\right)\right)$

and $K_2(c_\infty(1-\alpha))$. In practice, based on equation (2.3), the subsampling size is likely to be chosen as $b_n = Cn^{2/3}$ for a fixed $C > 0$. In this case, unless $K_1(c_\infty(1-\alpha))C^{-1/2} + K_2(c_\infty(1-\alpha))C = 0$, the subsampling approximation 1 will also converge to the distribution of interest at a rate of $n^{-1/3}$.

According to previous sections, the bootstrap delivers an error in the coverage probability of order $n^{-1/2}$. Hence, in the conditionally separable model and under certain conditions, the error in the coverage probability of the bootstrap approximation is eventually smaller than the error in the coverage probability produced by the subsampling approximation 1.

In the appendix, we show that the subsampling approximation 2 generates consistent inference in level (theorem A.13) and we provide an upper bound on the rate of convergence of the error in coverage probability (theorem A.14). In this case, the upper bound on the rate of convergence is of order $b_n^{-1/2} + \tau_n (b_n/n)^{1/2}$, which is even worse than the one obtained for subsampling 1. Moreover, lemma A.16 in the appendix provides conditions under which this rate is not just an upper bound, but the exact rate of convergence of the error in the coverage probability.

2.3.2 Asymptotic approximation

Theorem A.1 shows that the limiting distribution of the statistic of interest converges weakly to a continuous function of a tight Gaussian process with a certain variance covariance function. An asymptotic approximation can be constructed by replacing the Gaussian process with an estimate. This procedure is described in detail in section A.6.2 of the appendix.

Following the steps we used for the bootstrap approximation, we can prove consistency in level for the asymptotic approximation.

Theorem 2.5 *[Consistency in level - asymptotic approximation] Assume (A1)-(A4) and (CF). If the identified set is non-empty then, for any* $\alpha \in (0, 0.5)$,

$$\lim_{n\to+\infty} P\left(\Theta_I \subseteq \hat{C}_n^{AA}(1-\alpha)\right) = (1-\alpha)$$

Moreover, we can also establish the rate of convergence of the error in the coverage probability of the asymptotic approximation.

Theorem 2.6 *[ECP - asymptotic approximation] Assume (B1)-(B4) and (CF). If the identified set is non-empty then, for any* $\alpha \in (0, 0.5)$,

$$\left| P\left(\Theta_I \subseteq \hat{C}_n^{AA}(1-\alpha)\right) - (1-\alpha)\right| = O\left(n^{-1/2}\right)$$

For the conditionally separable model, the bootstrap and the asymptotic approximation have the same upper bound on the rate of convergence of the error in the coverage probability. In other words, the bootstrap does not seem to be providing any improvement over the asymptotic approximation, was is usually referred to as asymptotic refinements[23]. The lack of asymptotic refinements is expected given that the statistic of interest is not asymptotically pivotal[24]. Finally, notice that the implementation of the asymptotic approximation requires a simulation that involves exactly the same amount of computation as the implementation of the bootstrap approximation.

3 Conclusion

This paper contributes to the growing literature of inference in partially identified or set identified econometric models. We build on the criterion function approach proposed by CHT [10].

The first contribution of this paper is to introduce a novel bootstrap procedure to construct confidence sets that cover the identified set with a prespecified probability in a wide class of partially identified models. The models considered are those defined by finitely many moment inequalities and equalities, which include many applications of economic interest. Asymptotically, the coverage level provided by our confidence set converges to the desired coverage level, that is, our procedure is consistent in level. This constitutes an advantage relative to other inferential procedures that have been proposed in the literature. Along the lines of our bootstrap procedure, we also propose an asymptotic approximation that is also consistent in level[25].

The second contribution of our paper is to analyze the rate of convergence of the error in the coverage probability for our bootstrap approximation, our asymptotic approximation and for subsampling approximations such as the one proposed by CHT [10]. We show that our bootstrap approximation and our asymptotic approximation have errors in the coverage probability of order $n^{-1/2}$. Under certain conditions, we show that the error in the coverage probability of the subsampling approximation converges to zero at a rate of $n^{-1/3}$ or slower. As a consequence, under these conditions, our bootstrap and our asymptotic approximation should eventually provide inference that is more precise than that of the subsampling approximation.

The Monte Carlo simulations presented in the appendix reveal that our bootstrap approximation and our asymptotic approximation have a satisfactory finite sample performance. Moreover, both of these approximations exhibit a much better finite sample performance than the subsampling procedures, in accordance to the results regarding rates of convergence of the error in the coverage probability.

This paper opens several topics for further research. A first topic is to find an adequate data dependent procedure to choose the sequence $\{\tau_n\}_{n=1}^{+\infty}$, which is the key to our consistency result. A second important generalization of this paper would be to allow for continuous covariates, which

[23] In order to obtain asymptotic refinements, one could consider a computationally intensive procedure called prepivoting. This procedure is described in detail in Hall [14] and Horowitz [16]. The study of the validity of the prepivoting procedure in this setting is out the scope of this paper.

[24] This is discussed in section A.2.3 of the appendix.

[25] As we mentioned earlier, this approximation was independently introduced by Soares [28], Andrews and Soares [3], CHT [10] and working paper versions of this paper.

requires modifying the formal arguments in non trivial ways. A final topic left for future research is to study whether the theoretical properties of our inferential method (coverage in level and rates) hold uniformly over a relevant class of probability distributions. The literature that studies this problem in partially identified models has focused exclusively on the problem of coverage of each of the elements of the identified set with a prespecified probability. Extending these results to the problem of coverage of the identified set with a prespecified probability appears to be a very challenging problem.

References

[1] D.W.K. Andrews. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68(2):399–405, March 2000.

[2] D.W.K. Andrews, S. Berry, and P. Jia. Confidence regions for parameters in discrete games with multiple equilibria with an application to discount chain store location. Mimeo: Yale University and M.I.T., May 2004.

[3] D.W.K. Andrews and G. Soares. Inference for parameters defined by moment inequalities using generalized moment selection. Mimeo: Yale University, May 2007.

[4] G.J. Babu and K. Singh. Edgeworth expansions for sampling without replacement from finite populations. *Journal of Multivariate Analysis*, 17(3):261–278, December 1985.

[5] A. Beresteanu and F. Molinari. Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814, June 2008.

[6] R.N. Bhattacharya and R.R. Rao. *Normal Approximation and Asymptotic Expansions*. John Wiley and Sons, Inc., 1976.

[7] P. Bickel and R. Freedman. Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9(6):1196–1217, November 1981.

[8] F. A. Bugni. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the elements of the identified set. Mimeo: Duke University, April 2009.

[9] I. Canay. E. L. inference for partially identified models: large deviations optimality and bootstrap validity. Mimeo: Northwestern University, June 2009.

[10] V. Chernozhukov, H. Hong, and E. Tamer. Parameter set inference in a class of econometric models. *Econometrica*, 75(5):1243–1284, September 2007.

[11] Yu. A. Davydov, M. A. Lifshits, and N. V. Smorodina. *Local Properties of Distributions of Stochastic Functionals*, volume 173 of *Translations of Mathematical Monographs*. American Mathematical Society, 1995.

[12] R. Dudley. *Uniform Central Limit Theorems*. Cambridge University Press, 1999.

[13] E. Giné and J. Zinn. Bootstrapping general empirical measures. *Annals of Probability*, 18(2):851–869, April 1990.

[14] P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer, New York, 1992.

[15] J. H*ü*sler, R. Liu, and K. Singh. A formula for the tail probability of a multivariate normal distribution and its applications. *Journal of Multivariate Analysis*, 82(2):422–430, February 2002.

[16] J. L. Horowitz. *The Bootstrap*, volume 5 of *Handbook of Econometrics*, chapter 52, pages 3159–3228. Elsevier Science B.V., May 2002.

[17] G. Imbens and C.F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, November 2004.

[18] K. Joag-Dev and F. Proschan. Negative association of random variables with applications. *The Annals of Statistics*, 11(1):286–295, March 1983.

[19] J. Kuelbs. Kolmogorov's law of iterated logarithm for banach space valued random variables. *Illinois Journal of Mathematics*, 21(7):784–800, December 1977.

[20] C.F. Manski. Anatomy of the selection problem. *The Journal of Human Resources*, 24(3):343–360, Summer 1989.

[21] A. Pakes, J. Porter, K. Ho, and J. Ishii. Moment inequalities and their application. Mimeo: Harvard University, University of Wisconsin, Columbia University and, April 2006.

[22] D. N. Politis and J. P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22(4):2031–2050, December 1994.

[23] D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling*. Springer, 1999.

[24] J. P. Romano and A. M. Shaikh. Inference for the identified set in partially identified econometric models. Mimeo: Stanford University and University of Chicago, September 2006.

[25] J. P. Romano and A. M. Shaikh. Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference*, 138:2786–2807, 2008.

[26] A. Rosen. Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. Mimeo: University College London, June 2006.

[27] Q. Shao and C. Su. The law of the iterated logarithm for negatively associated random variables. *Stochastic Processes and their Applications*, 83:139–148, 1999.

[28] G. Soares. Inference for partially identified models with inequality moment constraints. Mimeo: Yale University, June 2006.

[29] A. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.