INFERENCE ON PARAMETER SETS IN ECONOMETRIC MODELS*

VICTOR CHERNOZHUKOV† HAN HONG§ ELIE TAMER‡

ABSTRACT. This paper provides confidence regions for minima of an econometric criterion function $Q(\theta)$. The minima form a *set* of parameters, Θ_I, called the identified set. In economic applications, Θ_I represents a class of economic models that are consistent with the data. Our inference procedures are criterion function based and so our confidence regions, which cover Θ_I with a prespecified probability, are appropriate level sets of $Q_n(\theta)$, the sample analog of $Q(\theta)$. When Θ_I is a singleton, our confidence sets reduce to the conventional confidence regions based on inverting the likelihood or other criterion functions. We show that our procedure is valid under general yet simple conditions, and we provide feasible resampling procedure for implementing the approach in practice. We then show that these general conditions hold in a wide class of parametric econometric models. In order to verify the conditions, we develop methods of analyzing the asymptotic behavior of econometric criterion functions under set identification and also characterize the rates of convergence of the confidence regions to the identified set. We apply our methods to regressions with interval data and set identified method of moments problems. We illustrate our methods in an empirical Monte Carlo study based on Current Population Survey data.

KEY WORDS: Set estimator, level sets, interval regression, subsampling bootsrap

Date: Original Version: May 2002. This version: June, 2004 .

* Note: This paper is currently undergoing MAJOR revision following referee comments. New results on estimation and rates of convergence are added. Convexity assumptions on the parameter space are being replaced by more general conditions. The most recent, yet preliminary and incomplete, revision can be requested via e-mail. September, 2005.

*We thank Alberto Abadie, Gary Chamberlain, Jinyong Hahn, Bo Honoré, Sergei Izmalkov, Guido Imbens, Christian Hansen, Yuichi Kitamura, Chuck Manski, Whitney Newey, and seminar participants at Berkeley, Brown, Chicago, Harvard-MIT, Montreal, Northwestern, Penn, Princeton, UCSD, Wisconsin, and participants at the Winter Meetings of Econometric Society in San-Diego for comments.

† Department of Economics, MIT, vchern@mit.edu. Research support from the National Science Foundation (SES-0241810) is gratefully acknowledged.

§ Department of Economics, Duke University, han.hong@duke.edu. Research support from the National Science Foundation (SES-0335113) is gratefully acknowledged.

‡ Department of Economics, Princeton University, tamer@princeton.edu. Research support from the National Science Foundation (SES-0112311) is gratefully acknowledged.

1. Introduction and Motivation

Parameters of interest in econometric models can be defined as those parameter vectors that minimize a population objective or criterion function. If this criterion function is minimized uniquely at a particular parameter vector, then one can obtain valid confidence regions (or intervals) for this parameter using a sample analog of this function. Likelihood and method of moments procedures are two commonly used and well studied methods in this setting. This paper extends this criterion-based inference to econometric models that are *set identified*, i.e., models where the objective function is minimized on a set of parameters, the *identified set*. Our goal is to make inferences directly on the identified set and to provide a method of obtaining confidence regions with good properties (such as consistency and equivariance) that cover the identified set with a prespecified probability.

Our point of departure is a nonnegative population criterion function $Q(\theta)$ and its finite sample analog $Q_n(\theta)$ where $\theta \in \Theta \subset \mathbb{R}^d$. The identified set can be defined as $\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$ where every θ in Θ_I indexes an economic model that is consistent with the data. The objective of this paper is to construct confidence sets C_n for Θ_I from the level sets of $Q_n(\theta)$ such that

$$\lim_{n\to\infty} P(\Theta_I \subseteq C_n) = \alpha, \tag{1.1}$$

for a prespecified confidence level $\alpha \in (0,1)$. A level set $C_n(c)$ of the finite sample criterion function Q_n is defined as

$$C_n(c) := \Big\{\theta : Q_n(\theta) - q_n \leq c/a_n\Big\}, \text{ where } q_n := \inf_{\theta\in\Theta} Q_n(\theta) \text{ or } q_n := 0,$$

for some appropriate normalization a_n. In order to obtain the correct coverage (1.1), we choose $C_n = C_n(c)$ with the cut-off level $c = c_\alpha$, where c_α equals the asymptotic α-quantile of the coverage statistic:

$$\mathcal{C}_n := \sup_{\theta\in\Theta_I} a_n(Q_n(\theta) - q_n),$$

which is a quasi-likelihood-ratio type quantity. The constructed confidence sets possess important properties, such as consistency and equivariance to reparameterization. Our approach covers general situations, where the identified set is defined as the minimand of an objective function. In addition, our confidence regions are sets that are robust to the failure of point-identifying assumptions in that they cover the unknown identified set – whether it is a set or a point – with a prespecified probability. These confidence sets collapse to the usual confidence intervals based on the likelihood ratio in cases where the identified set Θ_I is a singleton.

We show that the above level set $C_n(c_\alpha)$ has correct asymptotic coverage, where c_α is the appropriate quantile of a well defined coverage statistic $\mathcal{C}$, the nondegenerate large sample limit of $\mathcal{C}_n$. Then, we provide general resampling methods to consistently estimate the percentile c_α. The coverage and resampling results hold under general (high level) conditions. Focusing on a class of econometric models, we then show that these conditions are satisfied. In the process, we provide characterizations of the stochastic properties of the criterion functions process, $a_n(Q_n(\cdot) - q_n)$, exploiting the fact that the population criterion Q is minimized on a set rather than a point. Furthermore, we obtain the asymptotics of the coverage statistic $\mathcal{C}_n$ and of the level sets $C_n(c)$, focusing on coverage and speed of convergence of $C_n(c)$ to Θ_I. The paper then considers two important applications: regression with interval censored outcomes and set-identified generalized method-of-moments. We finally illustrate our methods in a Monte Carlo study based on data from the Current Population Survey.

In the last decade, a growing body of literature has considered the problem of inference in partially identified models, i.e., models where parameters of interest are set identified, cf. Manski (2003). While most of the work, e.g. Horowitz and Manski (1998) and Imbens and Manski (2004), exclusively focuses on the case where there is a scalar parameter of interest that lies in an interval, this paper is concerned with inference on vector parameters in problems where the identified set is defined via a general optimization problem, as in the economic problems described below. This multivariate set is usually not an interval (or cube) and can be a set of isolated points or manifolds. Moreover, the methods used in the interval case are based on estimating the two end-points of the interval. Hence they are not applicable in the general set-identified case, for which a different approach is needed.

The main motivation for posing the identified set as the object of inference is motivated by many examples. In Hansen, Heaton, and Luttmer (1995), the identified set is a subset of asset-pricing models that obey the pricing-error and volatility constraints implicit in asset market returns. In models with multiple equilibria, the identified set is the set of parameters that describe different equilibria supported across markets or industries. There is no single "true" equilibrium that is played, since particular equilibria may vary across observational units; see Ciliberto and Tamer (2003). Another example is the structural instrumental variable estimation of returns to schooling. Suppose that we are interested in the following example where potential income Y is related to education E through a flexible, quadratic functional form, $Y = \alpha_0 + \alpha_1 E + \alpha_2 E^2 + \epsilon$. Although parsimonious, this simple model is not point-identified in the presence of the standard quarter-of-birth instrument Z suggested in Angrist and Krueger (1992) (indicator of the first quarter of birth).[1] In the absence of point identification, all of the parameter values $(\alpha_0, \alpha_1, \alpha_2)$ consistent with the instrumental orthogonality restriction $E(Y - \alpha_0 + \alpha_1 E + \alpha_2 E^2)(1, Z)' = 0$ are of interest for

[1] In some cases the indicators of other quarters of birth are used as instruments. However, these instruments are not correlated with education (correlation is extremely small) and thus bring no additional identification information.

purposes of economic analysis. Similar partial identification problems arise in nonlinear moment and instrumental variables problems, see e.g. Demidenko (2000) and Chernozhukov and Hansen (2001).

Literature. To our knowledge, the earliest work in econometrics on parametric set identified models can be found in the paper of Marschak and Andrews (1944). There, the identified set is the collection of parameters representing different production functions that are consistent with the data and functional restrictions the authors consider.[2] Gisltein and Leamer (1983) provide set consistent estimation in a class of likelihood models where Θ_I as the set of parameters that are robust to misspecification. Klepper and Leamer (1984) generalize the Frish bounds to multivariate regression models with measurement errors. On the other hand, Hansen, Heaton, and Luttmer (1995) provide consistent set estimates of means and standard deviations in a class of asset pricing methods. Manski and Tamer (2002) provided conditions under which an appropriately defined set consistently estimates the identified set. However, these consistency results do not contain a method for inference about the identified set. To the best of our knowledge, there are no results in the literature that deal with the general problem of obtaining confidence regions for parameter sets.[3]

The remainder of the paper is organized as follows. Section 2 provides a general theory of inference on the identified sets based on general criterion functions. Section 3 focuses on a class of regular parametric models, provides verification of the regularity conditions posed in Section 2, and provides additional results that pertain to regular cases. Section 4 provides a Monte Carlo evaluation of the methods, and Section 5 concludes.

2. General Set Inference in Large Samples

2.1. **Generic Inference on Identified Sets.** In this section we present our main result which forms the basis for the rest of the analysis. We first define the identified set Θ_I and formalize some definitions that will be used throughout.

For given data, the inference about the parameter set Θ_I is based on a criterion function $Q_n(\theta) = Q_n(\theta, W_1, ..., W_n)$, where data $\{W_1, ..., W_n\}$ are defined on some common probability space (Ω, F, P). The criterion function $Q_n(\theta)$ converges to a continuous criterion function $Q(\theta)$, that is minimized at Θ_I, the identified set.

Assumption A.1 (Basic Setup)**.** *Criterions* $Q_n : \mathbb{R}^d \to \mathbb{R}^+$ *and* $Q : \mathbb{R}^d \to \mathbb{R}^+$ *and* Θ *satisfy*

[2]For a good description of Marschak and Andrews (1944), see Chapter III of Nerlove (1965).

[3]The only other paper known to us is Imbens and Manski (2004). Their paper considers a different problem of inference about the a real parameter θ^* that is interval-identified (i.e. contained between some upper and lower bounds that can be estimated.) The problem of inference about θ^* is fundamentally different from inference about Θ_I, as shown in Imbens and Manski (2004) and in Appendix G.

i. Θ *is a compact and convex subset of* $\mathbb{R}^d$ *[CONVEXITY to BE RELAXED],*

ii. $Q(\theta)$ *is continuous and* $Q_n(\theta)$ *is lower-semicontinuous,*

iii. $\Theta_I = \arg\min_{\theta\in\Theta}[Q(\theta)]$ *is a finite union of connected compact subsets of* Θ*,*

iv. $Q(\theta_I) = 0$ *for each* $\theta_I \in \Theta_I$*,*

v. $Q_n(\theta) - Q(\theta) = o_p(1)$ *for each* $\theta \in \Theta$*.*

Assumption A.1 states a standard compactness and convexity assumption, which are important to the subsequent analysis. It also defines Θ_I as the minimizer of the limit criterion function Q. The region Θ_I is a finite union of compact connected sets, an assumption that serves to organize the presentation. This covers both the case when Θ_I is a finite union of isolated points and the case when Θ_I is a finite union of compact sets with boundaries defined by manifolds (nonlinear hyperplanes).[4] The assumption that $Q(\theta_I) = 0$ serves as a convenient normalization. The pointwise convergence condition serves to relate Q as the limit of the finite sample objective function. The pointwise convergence will be strengthened later on.

Lemma 2.1 shows how to construct a level set of the sample objective function that will eventually provide the proper inferential statement (1.1) about Θ_I in a generic setting. The c-level set of objective function Q_n is given by

$$C_n(c) := \Big\{\theta : a_n\left(Q_n(\theta) - q_n\right) \le c\Big\}, \text{ where } q_n := \inf_{\theta\in\Theta} Q_n(\theta) \text{ or } q_n := 0,$$

where a_n is defined in Assumption A.2 below. Typically a_n equals n in the regular cases studied later. Note that choosing $q_n = \inf_{\theta\in\Theta} Q_n(\theta)$ guarantees that the confidence region $C_n(c)$ is always *non-empty*, though in some cases one has $\inf_{\theta\in\Theta} Q_n(\theta) = 0$ with probability converging to one, see e.g. Example 1 in Section 3.

Consider the following coverage index

$$\rho(c) = \quad c - \sup_{\theta\in\Theta_I} a_n\left(Q_n(\theta) - q_n\right). \tag{2.1}$$

The sign of the index $\rho(c)$ indicates whether $\Theta_I \subseteq C_n(c)$ or not. For example, if there is $\theta \in \Theta_I$ such that $\theta \notin C_n(c)$, we have $a_n\left(Q_n(\theta) - q_n\right) > c$ which implies that $\rho(c) < 0$, and vice versa. The index is also linear in c, which will allow us to have data-dependent cut-off levels c. Another desirable property is invariance of the index ρ to parameter transformations which implies equivariance of $C_n(c)$ to parameter transformations. For instance, a one-to-one transformation of parameters $\theta \to \tau(\theta)$ changes the level set in the equivariant way

$$\Big\{\tau(\theta) : a_n\left(Q_n(\tau^{-1}(\tau(\theta))) - q_n\right) \le c\Big\} = \tau\left(C_n(c)\right).$$

The following lemma summarizes the discussion.

---

[4]For definition of manifold, see e.g. Milnor (1964) .

Lemma 2.1. *Let Assumptions A.1-(ii) and A.1-(iii) hold. Then, I. the coverage property holds:* $\rho(c) < 0 \Leftrightarrow \Theta_I \not\subseteq C_n(c)$ *and* $\rho(c) \geq 0 \Leftrightarrow \Theta_I \subseteq C_n(c)$*; II. the coverage index is linear in the cut-off level; III. the coverage index is invariant to re-parameterization; and IV. the level sets are equivariant to bijective parameter transformations.*

Next we state the main condition that enables large sample inference on Θ_I.

Assumption A.2 (Coverage Statistic)**.** *Suppose that there exist a sequence of constants* $a_n \to \infty$ *such that*

$$\mathcal{C}_n = \sup_{\theta_I \in \Theta_I} a_n \left(Q_n(\theta_I) - q_n \right) \to_d \mathcal{C},$$

where $\mathcal{C}$ *is a nondegenerate random variable.*

In the sequel we explain how this main assumption is attained in sufficiently regular models of interest, where $a_n = n$. We also provide methods for verification of this assumption and for finding the limit variable $\mathcal{C}$. Verification of Assumption A.2 is a difficult matter and requires developing a set of new asymptotic methods of dealing with criterion functions under set identification.

Assumption A.2 leads us to the following main theorem, which provides a generic result on inference in set-identified models.

Theorem 2.1 (Generic Inference on Sets)**.** *Suppose Assumptions A.1 and A.2 hold and that* c_α *is a continuity point of distribution function of* $\mathcal{C}$ *such that* $P\{\mathcal{C} \leq c_\alpha\} = \alpha$*. Then for any* $\hat{c}_\alpha \to_p c_\alpha$,

$$I. \quad \rho\left(C_n(\hat{c}_\alpha)\right) \to_d c_\alpha - \mathcal{C} \text{ and } II. \quad \lim_{n \to \infty} P\Big\{\Theta_I \subseteq C_n(\hat{c}_\alpha)\Big\} = \alpha.$$

The result follows immediately from Lemma 2.1 and A.2. First, $\rho(C_n(c)) = \mathcal{C}_n - \hat{c}_\alpha = \mathcal{C}_n - c_\alpha + o_p(1) \to_d \mathcal{C} - c_\alpha$. Second,

$$\begin{aligned} P\Big\{\Theta_I \subseteq C_n(\hat{c}_\alpha)\Big\} = P\Big\{\rho(C_n(\hat{c}_\alpha)) \geq 0\Big\} &= P\Big\{\hat{c}_\alpha - \mathcal{C}_n \geq 0\Big\} \\ &= P\Big\{\mathcal{C}_n \leq c_\alpha + o_p(1)\Big\} = P\Big\{\mathcal{C} \leq c_\alpha\Big\} + o(1), \end{aligned}$$

provided c_α a continuity point of distribution function of $\mathcal{C}$.

It is clear from Theorem 2.1 that our method of constructing valid confidence sets builds on the classical principle of inverting some criterion function. Indeed, in point identified cases Θ_I reduces to a singleton $\{\theta_I\}$ so that the coverage statistic becomes a standard likelihood ratio type quantity $\mathcal{C}_n = a_n \left(Q_n(\theta_I) - q_n \right)$, which follows well known limit laws. In the general case, the statistic is a more involved quantity, being the supremum over the elements of Θ_I: $\mathcal{C}_n = \sup_{\theta_I \in \Theta_I} a_n \left(Q_n(\theta_I) - q_n \right)$.

In practice, our procedure for constructing the confidence regions critically depends on being able to consistently estimate c_α, the α-quantile of the limit variable $\mathcal{C}$. In all examples that we study,

$\mathcal{C}$ is non-standard and its distribution depends on Θ_I. Despite this problem, we will show how to obtain a consistent estimate of c_α by the following method.

2.2. **Feasible Inference.** We first need to obtain an approximation to the sampling distribution of $\mathcal{C}_n = \sup_{\theta \in \Theta_I} a_n \left(Q_n(\theta) - q_n \right)$. Since we do not observe Θ_I, we will replace it by an *initial estimate*[5]

$$\widehat{\Theta}_I = C_n(\hat{k}), \text{ where } \hat{k} \in [c_1, c_2] \cdot \ln n \quad \text{wp} \to 1, \tag{2.2}$$

where $\hat{k}$ is a possibly data-dependent *starting value* of the cut-off. (In the first class of our examples, we can use th starting cut-off $\hat{k} = 0$.) The result proven below suggests that the asymptotic validity of the procedure will not depend on the starting value. In finite samples, the choice of the starting value may be important, and we discuss it in Section 4.

Consider the following subsampling algorithm:

1. For cases when data $\{W_i\}$ are iid, construct all (B_n) subsets of size $b \ll n$ of the data. For cases when $\{W_t\}$ form a stationary time series, construct $B_n = n - b + 1$ subsets of size b of the form $\{W_i, ..., W_{i+b-1}\}$. (In practice, one can use a smaller number B_n of randomly chosen subsets under the condition that $B_n \to \infty$ as $n \to \infty$.)

2. For each $j = 1, ..., B_n$, compute

$$\widehat{\mathcal{C}}_{j,b,n} = \sup_{\theta \in C_n(\hat{c})} a_b \left(Q_{j,b}(\theta) - q_{j,b} \right),$$

where $q_{j,b} := 0$ if $q_n := 0$ and $q_{j,b} := \inf_{\theta \in \Theta} Q_{j,b}(\theta)$ otherwise; $Q_{j,b}(\theta)$ denotes the criterion function defined using the j-th subset of the data only.

3. Let $\hat{c}_\alpha$ be the α-th quantile of the sample $\{\widehat{\mathcal{C}}_{j,b,n}, j = 1, ..., B_n\}$.

4. (Optional) As commented in Section 4.3, one could repeat steps 2 and 3 finite number of times using $\hat{c} = \hat{c}_\alpha \ln n$.[6] (Any finite number of repetitions produces the consistent estimate of c_α.)[7]

We require that as $n \to \infty$,

$$b/n \to 0, \quad B_n \to \infty, \quad b \to \infty \quad \text{at polynomial rates.} \tag{2.3}$$

The choice of b and other practical aspects of the procedure are discussed in Section 4.

To guarantee asymptotic validity of the above procedure, the following assumption is needed.

[5]The adjustment factor $\ln n$ can be replaced by $\ln \ln n$ or any other $m(n)$ such that $m(n) \to \infty$ and $m(n)/n^a \to 0$ as $n \to \infty$ for all $a > 0$.

[6]The adjustment by $\ln n$ is not needed in the first class of our examples.

[7]In fact, we found in Monte-Carlo experiments described in Section 4 that just two repetitions worked the best in terms of coverage and computational expense. This was also confirmed by Bajari, Benkard, Levin (2003).

Assumption A.3 (Sandwich Property)**.** *Suppose that $b/n \to 0$ and $b \to \infty$ at polynomial rates as $n \to \infty$. For $\hat{k} \in [c_0, c_1] \cdot \ln n$, wp $\to$ 1*

$$\Theta_I \subset C_n(\hat{k}) \subset \Theta_I^{\epsilon_n}, \text{ such that } \quad a_b(\sup_{\theta \in \Theta_I^{\epsilon_n}} Q_b(\theta) - \sup_{\theta \in \Theta_I} Q_b(\theta)) = o_p(1),$$

$\Theta_I^{\epsilon_n} := \{\theta_I + t : \|t\| \leq \epsilon_n, \theta_I \in \Theta_I\}$ and $\epsilon_n \to 0$ is a sequence of positive constants.

Assumption A.3 guarantees *inferential validity* but also allows us to establish consistency and to characterize the *rate of convergence* of the level sets $C_n(\hat{k})$ to the identified set Θ_I, as shown below in Theorem 2.2. The intuition behind A.3 is as follows. In the subsampling bootstrap, we do not know Θ_I, hence we replace it by an estimate $C_n(\hat{c})$. The replacement should have only a negligible impact on the distribution of the coverage statistic in subsamples. A.3 is similar in nature to the polynomial rate of convergence assumption used by Politis, Romano, and Wolf (1999), p 44, in the case of Wald inference in point identified cases. There, one does not know the true θ_I and replaces it with an estimate $\hat{\theta}_I$, hence requiring that this replacement has negligible effect on the distribution of the Wald statistic in subsamples. We show how to verify A.3 for parametric models in Section 3 (Theorem 3.2). The next theorem summarizes our results for general inference in set identified models.

Theorem 2.2 (Consistency and Validity of Inference)**.** *Suppose that the estimation data $\{W_i, i \leq n\}$ are iid or form a stationary strongly-mixing sequence and that A.2 and A.3 hold.*

I. *Then for the subsampling algorithm defined above, provided $P\{\mathcal{C} \leq c\}$ is continuous at $c = c_\alpha$, $\hat{c}_\alpha \to c_\alpha$, and*

$$P\left(\Theta_I \in C_n(\hat{c}_\alpha)\right) \to \alpha.$$

II. *The set $C_n(\hat{k})$ for $\hat{k} = \hat{c}_\alpha \ln n$ is consistent under the Hausdorff metric: wp $\to$ 1*

$$d_H(C_n(\hat{k}), \Theta_I) \leq \epsilon_n \to 0.$$

Recall that the Hausdorff metric between two sets is defined as:

$$d_H(A, B) := \max[h(A, B), h(B, A)], \quad \text{where} \quad h(A, B) := \sup_{a \in A} \inf_{b \in B} \|a - b\|.$$

Thus, under general conditions our inference method is asymptotically valid and delivers level sets $C_n(\hat{k})$ that converge to the identified set at the rate ϵ_n with respect to the Hausdorff distance (the rate ϵ_n will be shown to be essentially $1/\sqrt{n}$ for parametric models).

Under the given level of generality, subsampling appears to be the only valid resampling method. The conventional (n-out-of-n) bootstrap will not be generically consistent in the present settings. One counterexample is as follows: In Section 3, one of the leading examples, the partially identified linear IV regression, necessarily involves a parameter on the boundary problem. It is known that the bootstrap fails in parameter-on-the boundary problems, cf. Andrews (2000).

3. Regular Parametric Case and Applications

In this section we establish the methods for verifying the main conditions required for implementing the approach, namely that existence of the limit distribution for the coverage statistic (A.2) and the sandwich condition (A.3). We develop these methods to cover a variety of parametric examples, and illustrate the the approach with applications to regression models with missing outcome data and generalized method of moments under partial identification. For parametric models, we establish further properties of the confidence regions, such as the speed of convergence of the level sets to the identified set, and the stochastic properties of objective functions in partially identified cases.

3.1. **Examples of Parametric Problems with Set Identification. Example I. Regression with Interval-Censored Outcomes.**

Consider the linear conditional expectation models

$$E_{P_\theta}[Y|X] = X'\theta, \text{ where } \theta \in \Theta, X \in \mathbb{R}^d.$$

The models are not assumed to be correctly specified, in the sense that there may be no θ such that $E_{P_\theta}[Y|X]$ agrees with $E_P[Y|X]$ under the actual law P of the data.

Observed data consists of i.i.d. observations (Y_{1i}, Y_{2i}, X_i) where Y_{1i} and Y_{2i} represents the interval observation on Y_i:

$$Y_i \in [Y_{1i}, Y_{2i}] \text{ given } X_i, \text{ a.s.} \tag{3.1}$$

In the absence of further information, the set

$$\Theta_I \equiv \{\theta \in \mathbb{R}^d : E[Y_1|X] \leq X'\theta \leq E[Y_2|X] \text{ a.s. } \} \tag{3.2}$$

is the object of interest, as it represents the set of linear conditional expectation models that are consistent with data. We assume that $\Theta_I \subset \Theta$, where Θ is a compact subset of $\mathbb{R}^d$. Observe that Θ_I minimizes the objective function

$$Q(\theta) = \int \left\{ \left(E[Y_1|x] - x'\theta\right)_+^2 + \left(E[Y_2|x] - x'\theta\right)_-^2 \right\} dP(x), \tag{3.3}$$

where $(u)_+^2 = (u)^2 \times 1[u > 0]$ and $(u)_-^2 = (u)^2 \times 1[u < 0]$. Notice also that $\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$. Using a sample analog of (3.3),

$$Q_n(\theta) = \frac{1}{n} \sum_{i \leq n} \left(\hat{E}[Y_{1i}|X_i] - X_i'\theta\right)_+^2 + \left(\hat{E}[Y_{2i}|X_i] - X_i'\theta\right)_-^2, \tag{3.4}$$

Manski and Tamer (2002) characterize consistent estimates of Θ_I.[8] However, no method for inference about Θ_I in the sense of providing the inferential statements was given. We provide below a confidence approach to inference about Θ_I in this and similar examples.

Example II. Structural Moment Equations. In method of moments settings, we are interested in deducing the set of all economic models indexed by a parameter $\theta \in \mathbb{R}^d$ that satisfy the moment equation computed with respect to the probability sampling distribution of the economic data. The data $(X_i, i \leq n)$ are stationary and strongly mixing, defined on the probability space $(\Omega, \mathcal{F}, P)$. The economic models θ_I of interest are assumed to satisfy

$$E[m_i(\theta_I)] = 0, \tag{3.5}$$

where $m_i(\theta) = m(\theta, X_i)$ is a lower-semi-continuous function in θ a.s. The entire set of models $\Theta_I \subset \Theta$ that solve (3.5) also minimize the criterion function

$$Q(\theta) = E[m_i(\theta)]' W(\theta) E[m_i(\theta)], \tag{3.6}$$

where $Q(\theta)$ is continuous for each $\theta \in \Theta$, a compact subset of $\mathbb{R}^d$, and $W(\theta)$ is a continuous and positive definite matrix for each $\theta \in \Theta$. In nonlinear models, the existence of multiple solutions to nonlinear equations (3.5) is more of a rule rather than an exception, except in special cases, see e.g. Demidenko (2000). In linear models there may be multiple solutions as well if the usual rank conditions fail to hold, as mentioned in the introduction. Thus, the GMM function (3.6) will in general be minimized on a set. The inference on Θ_I may be based on the usual GMM function

$$Q_n(\theta) = n[g_n(\theta)]' W_n(\theta)[g_n(\theta)], \quad g_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m_i(\theta), \tag{3.7}$$

where $W_n(\theta)$ is a lower-semi-continuous, uniformly positive definite matrix.[9] To the best of our knowledge, no methods for inference about Θ_I in the sense of providing Neyman-Pearson confidence statements has been yet given in such settings.

3.2. **General Asymptotics of Criterion Functions in Parametric Models.** The prime goal of this section is to provide primitive conditions and tools that help verify conditions A.2 and A.3 in a wide variety of parametric models. The tools will be illustrated using the examples stated in the next section.

[8]Specifically, they show that the set $\Theta_n = \{\theta \in \Theta : Q_n(\theta) \leq \min_\theta Q_n(\theta) + \epsilon_n\}$ converges almost surely to the set Θ_I if $\epsilon_n \geq 0$ converges to zero at a rate strictly slower than the rate at which $Q_n(\cdot)$ converges uniformly to $Q(\cdot)$. They use a Hausdorff metric as a distance function between sets.

[9]For example, it is convenient to use the continuous updating type weight matrix $W_n(\theta)$, i.e. to let $W_n(\theta)$ equal a consistent estimate of asymptotic variance of $n^{-1/2}\sum_{i=1}^n m_i(\theta)$.

Since $Q_n(\theta)$ approaches $Q(\theta)$, we should be able to tell that $\theta \notin \Theta_I$ if θ is outside some neighborhood of Θ_I. The size of this neighborhood depends on the rate at which the boundary of Θ_I can be learned. In the remainder of the paper, the rate we consider is the parametric rate and hence the points of uncertainty are those θ_I that are within a $1/\sqrt{n}$ neighborhood of the true set Θ_I. Such points are of the form

$$\theta_I + \lambda/\sqrt{n}, \text{ for } \theta_I \in \Theta_I, \lambda \in \mathbb{R}^d.$$

In order to keep track of such points it will suffice to record all pairs of the form (θ_I, λ). Hence of central interest is the local empirical process

$$(\theta_I, \lambda) \mapsto \ell_n(\theta_I, \lambda) := n\Big(Q_n(\theta_I + \lambda/\sqrt{n}) - Q(\theta_I)\Big),$$

over a suitable domain. The pertinent domain for θ_I will be shown to be the boundary of identified set $\partial\Theta_I$. Given $\theta_I \in \partial\Theta_I$, the pertinent domain for λ is

$$V_n(\theta_I) := \{\lambda \in \mathbb{R}^d : \theta_I + \lambda/\sqrt{n'} \in \Theta, \text{ for all } n' \in [n, \infty)\}.$$

In addition, the following limit version of $V_n(\theta_I)$ will play an important role:

$$\begin{aligned} &V_\infty(\theta_I) := \{\lambda \in \mathbb{R}^d : \theta_I + \lambda/\sqrt{n} \in \Theta \text{ for all sufficiently large } n\}, \\ &\text{where when } \theta_I \in \text{int}(\Theta), \quad V_\infty(\theta_I) = \mathbb{R}^d. \end{aligned} \tag{3.8}$$

Thus, $V_\infty(\theta_I)$ plays the role of the limit local parameter space relative to θ_I. When θ_I is in the interior of the parameter space, i.e. $\theta_I \in \text{int}(\Theta)$, $V_\infty(\theta_I) = \mathbb{R}^d$. When θ_I is on the boundary of the parameter space, i.e. $\theta_I \in \partial\Theta$, the local deviations λ should be constrained to the local parameter space $V_\infty(\theta_I)$ of the specified form. This situation is similar to the one arising in the point-identified case, as characterized in Andrews (1999). Unlike in the point-identified case, the boundary problem in the partially identified models is more of a rule rather than an exception. It arises even in the simplest leading cases, as will be seen in Section 3.3.

Another important set is the subset of the local parameter space $V_n(\theta_I)$ where the local deviations λ are constrained to be towards the interior of the identified set:

$$\Lambda_n(\theta_I) := \{0\} \cup \{\lambda \in \mathbb{R}^d : \theta_I + \lambda/\sqrt{n'} \in \text{int}(\Theta_I) \text{ for all } n' \in [n, \infty)\}.$$

In addition, the following limit version of $\Lambda_n(\theta_I)$ will play an important role:

$$\Lambda_\infty(\theta_I) := \{0\} \cup \{\lambda \in \mathbb{R}^d : \theta_I + \lambda/\sqrt{n} \in \text{int}(\Theta_I) \;\text{ for all sufficiently large }\; n\}.$$

Note that since $\text{int}(\Theta_I) \subset \Theta$, $\Lambda_n(\theta_I)$ is a subset of $V_n(\theta_I)$, which is a subset of $V_\infty(\theta_I)$, and $\Lambda_\infty(\theta_I)$ is also a subset of $V_\infty(\theta_I)$.

Given above definitions, we shall obtain the following representations of the coverage statistic

$$
\begin{aligned}
\mathcal{C}_n &:= \sup_{\theta_I \in \Theta_I} n\left(Q_n(\theta_I) - q_n\right) \\
&= \begin{cases} \sup_{\theta_I \in \Theta_I} \ \ell_n(\theta_I, 0) & \text{if} \quad q_n := 0, \\ \sup_{\theta_I \in \Theta_I} \ \ell_n(\theta_I, 0) - \inf_{\theta_I + \lambda/\sqrt{n} \in \Theta} \ \ell_n(\theta_I, \lambda) & \text{if} \quad q_n := \inf_{\theta \in \Theta} Q_n(\theta), \end{cases} \\
&= \begin{cases} \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_n(\theta_I)} \ \ell_n(\theta_I, \lambda) + o_p(1) & \text{if } nq_n \to_p 0, \\ \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_n(\theta_I)} \ \ell_n(\theta_I, \lambda) - \inf_{\theta_I \in \Theta_I, \lambda \in V_n(\theta_I)} \ \ell_n(\theta_I, \lambda) + o_p(1) & \text{if } nq_n \not\to_p 0. \end{cases}
\end{aligned}
$$

In this representation, the first equality follows by definition of the local empirical process $\ell_n(\theta, \lambda)$, while the second equality emerges from the analysis given in the appendix. To guarantee non-degenerate asymptotics for this statistic, we will require that $\ell_n(\theta_I, \lambda)$ converges to a sufficiently well-behaved random element $\ell_\infty(\theta_I, \lambda)$, in the finite-dimensional sense coupled with some uniformity, which we will call the *quasi-uniform convergence*. This form of convergence will be sufficient to establish that

$$
\mathcal{C}_n \to_d \ \mathcal{C} := \begin{cases} \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ \ell_\infty(\theta_I, \lambda), & \text{if } nq_n \to_p 0, \\ \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ \ell_\infty(\theta_I, \lambda) - \inf_{\theta_I \in \Theta_I, \lambda \in V_\infty(\theta_I)} \ \ell_\infty(\theta_I, \lambda) & \text{if } nq_n \not\to_p 0. \end{cases}
$$

The following conditions ensure that this convergence takes place.

Assumption C.1 (Degenerate Asymptotics on the Interior)**.** *When* $\text{int}(\Theta_I)$ *is not empty, so that* $\Theta_I \neq \partial\Theta_I$*, for any* $\epsilon > 0$ *there exists large enough* $\delta > 0$ *and large enough* n *such that for* $I_n(\delta) := \{\theta_I \in \text{int}(\Theta_I) : \inf_{\theta'_I \in \partial\Theta_I} \|\theta_I - \theta'_I\| > \delta/\sqrt{n}\}$, $\sup_{\theta_I \in I_n(\delta)} |\ell_n(\theta_I, 0)| = 0$ *with probability no less than* $1 - \epsilon$*. Since* $\ell_n(\theta_I, \lambda) \geq 0$*, this implies that*

$$
nq_n = \inf_{\theta_I \in \Theta_I, \lambda \in V_n(\theta_I)} \ell_n(\theta_I, \lambda) = 0 \ \ wp \to 1.
$$

Condition C.1 is motivated by the fact that in interval regression examples for large n

$$
\ell_n(\theta_I, \lambda) = 0 \text{ wp} \to 1 \text{ for any fixed } \theta_I \in \text{int}(\Theta_I) \text{ and } \lambda \in \mathbb{R}^d.
$$

We provide examples and further discussion of this interesting phenomenon in Section 3.3.

Condition C.2 puts further convergence conditions by appropriately matching the large sample behavior of function $\ell_n(\theta, \lambda)$ with that of some function $\ell_\infty(\theta, \lambda)$, which we call the *quasi-uniform limit* of $\ell_n(\theta_I, \lambda)$. In order to state the condition, define the following key functionals

$$
u_n(\lambda) := \sup_{\theta_I \in \partial\Theta_I} \ell_n(\theta_I, \lambda) \text{ and } l_n(\lambda) := \inf_{\theta_I \in \partial\Theta_I} \ell_n(\theta_I, \lambda) \text{ for } n < \infty \text{ and } n = \infty.
$$

Assumption C.2 (Quasi-Uniform Convergence Near Boundary)**.** *Let* $\theta_I \in \partial\Theta_I$ *and* $\lambda \in K$*, where* K *is any compact subset of* $\mathbb{R}^d$*. Then*

i. $\ell_n(\theta_I, \lambda) \geq 0$ *converges weakly in finite-dimensional sense to some function* $\ell_\infty(\theta_I, \lambda) \geq 0$, *which is continuous in* λ *for each* θ_I.

ii. *(a) if* $\Theta_I \neq \partial\Theta_I$, $u_n(\lambda)$ *converges weakly to* $u_\infty(\lambda)$ *in finite-dimensional sense, and (b)* $u_n(\lambda)$ *is stochastically equi-continuous; if* $\Theta_I = \partial\Theta_I$, *then (a)* $(u_n(\lambda), l_n(\lambda))$ *jointly converge weakly to* $(u_\infty(\lambda), l_\infty(\lambda))$ *in finite-dimensional sense, and (b)* $(u_n(\lambda)$ *and* $l_n(\lambda)$ *are stochastically equi-continuous.*

iii. $\sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) < \infty$ *a.s.*

Condition C.2 extends the standard conditions used to derive the asymptotics of criterion functions in the point-identified case, where Θ_I is a single point θ_I, see e.g. Newey and McFadden (1994). Although asymptotics in the interior is degenerate, as condition C.1 states, the asymptotics near the boundary is assumed to be non-degenerate. C.2-(i.-ii.) imposes the conditions required for obtaining the limit distribution of coverage statistics and for verifying A.3. C.2-(i) requires that $\ell_n(\theta, \lambda)$ converges in finite-dimensional sense to some limit $\ell_\infty(\theta, \lambda)$. C.2-(ii) requires that sup and inf transformations of $\ell_n(\theta_I, \lambda)$ over $\partial\Theta_I$, denoted $u_n(\lambda)$ and $l_n(\lambda)$, converge in finite-dimensional sense to respective sup and inf transformations of $\ell_\infty(\theta_I, \lambda)$, denoted as $u_\infty(\lambda)$ and $l_\infty(\lambda)$. It also requires that $u_n(\lambda)$ and $l_n(\lambda)$ are stochastically equicontinuous. C.2-(iii) insures tightness of the limit coverage statistic $\mathcal{C}$, since $\sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda)$ will be one of the components of $\mathcal{C}$.

Note that in C.2-(ii), it is possible to require that

$$(\theta_I, \lambda) \mapsto \ell_n(\theta_I, \lambda) \text{ is stochastically equicontinuous over } \partial\Theta_I \times K,$$

where K is any compact subsets of $\mathbb{R}^d$. This stronger and simpler condition certainly implies C.2-(ii). However, this stronger condition fails to hold in some cases of interest (the regression model with interval-censored outcome), while it holds in others (the generalized method of moments).

Assumption C.3 (Local Quadratic Bound). *For any* $\epsilon > 0$, *there is a sufficiently large positive* K *such that with probability at least as large as* $1 - \epsilon$ *for large enough* n

$$\ell_n(\theta_I, \lambda) \geq C_1 \cdot n \cdot \min[\nu(\theta_I, \lambda), \delta]^2$$

uniformly in (θ_I, λ) *such that* $\nu(\theta_I, \lambda) \geq K/\sqrt{n}$, *where* $\nu(\theta_I, \lambda) = \inf_{\theta_I' \in \Theta_I} \|\theta_I + \lambda/\sqrt{n} - \theta_I'\|$, *for some positive constants* C_1 *and* δ *that do not depend on* ϵ.

Condition C.3 is needed to obtain rate of convergence for set estimates, and along with C.1 and C.2 allows us to verify the main previous high-level conditions A.2 and A.3. C.3 extends, to the set-identified case, the standard conditions needed to obtain the rate of convergence of extremum estimators in point-identified case; see conditions in Theorem 5.52 in Van der Vaart (1998) and Newey and McFadden (1994), p. 2185.

Theorem 3.1 (Limits of $\mathcal{C}_n$). *Suppose A.1 and C.1-C.3 hold. Then A.2 holds, in particular,*

$$\mathcal{C}_n \to_d \mathcal{C} := \begin{cases} \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda), & \text{if } nq_n \to_p 0 \\ \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) - \inf_{\theta_I \in \Theta_I, \lambda \in V_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda), & \text{if } nq_n \not\to_p 0 \end{cases}$$

where $nq_n \to_p 0$ necessarily occurs if $q_n := 0$ or if $\text{int}(\Theta_I)$ is non-empty in Θ.

Theorem 3.1 verifies Assumption A.2. The theorem states that under a set of conditions, which may be easily checked in examples of interest, the coverage statistic attains a limit distribution, which is determined by the appropriate inf and sup transformations of the quasi-uniform limit $\ell_\infty(\theta_I, \lambda)$ over $\partial\Theta_I$ and the local parameter spaces $\Lambda_\infty(\theta_I)$ and $V_\infty(\theta_I)$.

In addition to this basic result, we would like to know certain properties of the level sets.

Theorem 3.2 (Rate of Convergence and Sandwich Property). *Suppose A.1 and C.1 - C.3 hold. Then, for some positive constants C and C',*

I. *For any $\hat{k} \to_p \infty$ such that $\hat{k} = o_p(n)$,*

$$\Theta_I \subset C_n(\hat{k}) \subset \Theta_I^{\epsilon_n} \ wp \ \to 1, \ \textit{where } \epsilon_n = C \cdot \sqrt{\hat{k}/n},$$

so that

$$d_H(C_n(\hat{k}), \Theta_I) \leq C \cdot \sqrt{\hat{k}/n} \ wp \ \to 1;$$

II. *For any $\hat{k} \in [c_0, c_1] \ln n$, the sandwich condition A.3 takes place,*

$$\Theta_I \subset C_n(\hat{k}) \subset \Theta_I^{\epsilon_n} \ wp \ \to 1, \ \textit{where } \epsilon_n = C' \cdot \ln n / \sqrt{n},$$

and, provided $b \to \infty$ and $b/n \to 0$ at polynomial rate,

$$b(\sup_{\theta \in \Theta_I^{\epsilon_n}} Q_b(\theta) - \sup_{\theta \in \Theta_I} Q_b(\theta)) = o_p(1).$$

Theorem 3.2 verifies Assumption A.3 for parametric models and characterizes the rate of convergence of the level sets to the identified set. The rate of convergence, as measured by Hausdorff metric, is essentially $1/\sqrt{n}$.

3.3. **Analysis of Regression with Interval-Censored Outcomes.** First consider the case when $X_i = 1$ and suppose $E[Y_1] < E[Y_2]$. Then $\Theta_I = \{\theta : E[Y_1] \leq \theta \leq E[Y_2]\}$, that is $\Theta_I = [E[Y_1], E[Y_2]]$. Then

$$Q_n(\theta) = \left(\bar{Y}_1 - \theta\right)_+^2 + \left(\bar{Y}_2 - \theta\right)_-^2 ,$$

and

$$\ell_n(\theta_I, \lambda) = n\left((\bar{Y}_1 - \theta_I - \lambda/\sqrt{n})_+^2 + (\bar{Y}_2 - \theta_I - \lambda/\sqrt{n})_-^2\right).$$

Suppose that

$$\sqrt{n}(\bar{Y}_1 - EY_1, \bar{Y}_2 - EY_2)' \to_d (W_1, W_2)' \sim \mathcal{N}(0, \Omega).$$

Then

$$\ell_n(\theta_I, \lambda) = 0 \text{ wp} \to 1 \text{ if } \theta_I \in (EY_1, EY_2),$$

and the finite-dimensional limit of

$$(\ell_n(EY_1, \lambda),\ \ell_n(EY_2, \lambda))' \quad \text{is} \quad ((W_1 - \lambda)_+^2, (W_2 - \lambda)_-^2)'.$$

Therefore the finite-dimensional limit of $\ell_n(\theta, \lambda)$ is given by

$$\ell_\infty(\theta_I, \lambda) = (W_1 - \lambda)_+^2 1(\theta_I = EY_1) + (W_2 - \lambda)_-^2 1(\theta_I = EY_2).$$

Theorem 3.3 verifies C.1-C.3 for this example so that $\ell_\infty(\theta_I, \lambda)$ is also the quasi-uniform limit of $\ell_n(\theta_I, \lambda)$. This implies by Theorem 3.1 that

$$\mathcal{C}_n \to_d \mathcal{C} = \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) = \max\left[(W_1)_+^2,\ (W_2)_-^2\right].$$

One can use the above distribution to obtain the critical values and corresponding level set with the required coverage property.

Before proceeding to a more general case, notice that the inferential strategy developed in this simple example can be used in other problems where Θ_I is defined as the set of all $\theta \in \Theta$ such that $F_1 \leq \theta \leq F_2$ where F_1 and F_2 are functionals of the distribution of the observed data. This is a set of models that provide *interval bounds* on the parameters of interest. All one needs to do in this class of models is to derive the joint asymptotic distribution of the sample analogs of the endpoints of the interval of interest and obtain via simulation the critical values for the maximum of two random variables, as outlined above for the case when $F_1 = EY_1$ and $F_2 = EY_2$.

Getting back to more general settings, suppose that

$$X \in \{x_1, ..., x_J\} \quad P\text{-a.s.,} \tag{3.9}$$

where the first component of X is 1. Condition (3.9) assumes that X is discrete. The identified set Θ_I is determined by the set of inequalities:

$$\Theta_I := \{\theta \in \mathbb{R}^d : x_j'\theta \geq \tau_1(x_j) \text{ and } x_j'\theta \leq \tau_2(x_j) \text{ for all } j \leq J\}. \tag{3.10}$$

It is assumed that $\Theta_I \subset \Theta$, where Θ is a compact subset of $\mathbb{R}^d$, and that the $d \times J$ matrix

$$\mathcal{X} := (x_j, j \leq J) \text{ has rank } d, \tag{3.11}$$

which rules out redundant parameterization. Then the boundary of the identified set $\partial\Theta_I$ is

$$\partial\Theta_I = \{\theta_I \in \Theta_I : x_j'\theta_I = \tau_1(x_j) \text{ or } x_j'\theta_I = \tau_2(x_j), \text{ for some } j \leq J\}, \tag{3.12}$$

and $\text{int}(\Theta_I)$ is empty in $\mathbb{R}^d$, so that $\Theta_I = \partial\Theta_I$ if and only if $\tau_1(x_j) = \tau_2(x_j)$ for some j.

Define $\tau_1(x) = E(Y_1|x)$ and $\tau_2(x) = E(Y_2|x)$ and consider the objective function

$$\begin{aligned} Q_n(\theta) &:= \frac{1}{n}\sum_{i=1}^{n} \left(\hat{\tau}_1(X_i) - X_i'\theta\right)_+^2 + \left(X_i'\theta - \hat{\tau}_2(X_i)\right)_-^2, \\ \tau_k(x_j) &:= \sum_{i:X_i=x_j} Y_i, \quad k = 1, 2, \quad j = 1, ..., J. \end{aligned} \tag{3.13}$$

In this case for $n_j := \frac{1}{n}\sum_{i=1}^{n} 1[X_i = x_j]$,

$$\ell_n(\theta_I, \lambda) = n\sum_{j=1}^{J} \frac{n_j}{n}\left((\hat{\tau}_1(x_j) - x_j'\theta_I - x_j'\lambda/\sqrt{n})_+^2 + (\hat{\tau}_2(x_j) - x_j'\theta_I - x_j'\lambda/\sqrt{n})_-^2\right).$$

Define $\widehat{W}_{1j} := \sqrt{n}(\hat{\tau}_1(x_j) - \tau_1(x_j))$ and $\widehat{W}_{2j} := \sqrt{n}(\hat{\tau}_2(x_j) - \tau_2(x_j))$ for $j = 1, ..., J$. Assume also that a central limit theorem and a law of large numbers apply so that

$$\begin{aligned} &\left((\widehat{W}_{11}, \widehat{W}_{21}), ..., (\widehat{W}_{1J}, \widehat{W}_{2J})\right)' \to_d \left((W_{11}, W_{21}), ..., (W_{1J}, W_{2J})\right)' \sim \mathcal{N}(0, \Omega), \text{ and} \\ &n_j/n \to_p p_j \quad \text{for each } j = 1, ..., J. \end{aligned} \tag{3.14}$$

Theorem 3.3 (Interval Regression)**.** *Let the basic conditions specified in Section 3.2 hold and also (3.9)- (3.14) hold. Then C.1-C.3 and A.1-A.3 are satisfied. Moreover,*

$$\ell_\infty(\theta_I, \lambda) = \sum_{j=1}^{J} p_j\Big[(W_{1j} - x_j'\lambda)_+^2 1(x_j'\theta_I = \tau_1(x_j)) + (W_{2j} - x_j'\lambda)_-^2 1(x_j'\theta_I = \tau_2(x_j))\Big],$$

$$\mathcal{C} = \begin{cases} \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) & \text{if } nq_n \to_p 0 \\ \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) - \inf_{\theta_I \in \Theta_I, \lambda \in V_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) & \text{if } nq_n \not\to_p 0 \end{cases},$$

where $nq_n = 0$ *wp* $\to 1$ *occurs when* $\text{int}(\Theta_I)$ *is non-empty or if* $q_n := 0$.

This theorem verifies Assumptions C.1-C.3 and A.1-A.3, which implies that the results of Theorems 2.1, 2.2, 3.1, and 3.2 apply to the interval regression case. Therefore the inference procedure proposed in Section 2 is valid in this case. Further, the confidence regions have the stochastic properties given in Theorems 3.1 and 3.2.

The distribution of the limit variable $\mathcal{C}$ is not pivotal, and has no known closed analytical form, but this does not create a problem for the inferential method proposed in Section 2. $\mathcal{C}$ can not simplified much further, unlike in the trivial case with $X_i = 1$. One can simplify the leading term as

$$\sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) = \sup_{\theta_I \in \partial\Theta_I} \sum_{j=1}^{J} p_j \Big[(W_{1j})_+^2 1(x_j'\theta_I = \tau_1(x_j)) + (W_{2j})_-^2 1(x_j'\theta_I = \tau_2(x_j)) \Big],$$

but other terms do not appear to simplify in the above expressions.

3.4. **Analysis of the Structural Moment Equations Model.** We first work out a simple example, and then generalize it. Consider the usual two-stage-least-squares model

$$Y_i \equiv X_i'\theta_I + \epsilon_i(\theta_I), \quad E\epsilon_i(\theta_I)Z_i = 0,$$

which leads to the following objective function

$$Q_n(\theta) = \Big(\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i'\theta)Z_i'\Big) \Big(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\Big)^{-1} \Big(\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i'\theta)Z_i\Big).$$

Assume that $0 < r = \text{rank } EXZ' < \dim(\theta)$, so that the identified set Θ_I consists of an r-dimensional linear subspace of $\mathbb{R}^d$ intersected with Θ. Θ_I is defined as follows: for any θ_0 such that $E\epsilon(\theta_0)Z = 0$,

$$\Theta_I = \{\theta = \theta_0 + \delta \in \Theta \text{ such that } \delta' EXZ' = 0\}.$$

Note that Θ_I has empty interior relative to $\mathbb{R}^d$. Under the usual circumstances, even in the case when there is weak identification, $Q_n(\theta)$ is pivotal and can be inverted for confidence intervals. In the partially identified case, the statistic most relevant to making inference on Θ_I is the empirical process $(Q_n(\theta_I), \theta_I \in \Theta_I)$ which is no longer pivotal.

Suppose for simplicity that $q_n := 0$ (which is also true when $q_n := \inf_{\theta\in\Theta} Q_n(\theta)$ but $\dim(Z) \leq \dim(X)$), so that

$$\ell_n(\theta_I, \lambda) = \Big(\Delta_n(\theta_I) + \lambda'\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\Big)' \Big(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\Big)^{-1} \Big(\Delta_n(\theta_I) + \lambda'\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\Big),$$

$$\Delta_n(\theta_I) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i - X_i'\theta_I)X_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i(\theta_I)Z_i.$$

Under standard sampling conditions

$$\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \to_p EZZ', \quad \frac{1}{n}\sum_{i=1}^{n} Z_i X_i' \to_p EZX', \text{ and } \{\epsilon(\theta_I)Z, \theta_I \in \Theta_I\} \text{ is Donsker.}$$

Hence the finite-dimensional and quasi-uniform limit of $\ell_n(\theta_I, \lambda)$ is given by

$$\ell_\infty(\theta_I, \lambda) = \Big(\Delta(\theta_I) + \lambda' EXZ\Big)' \Big(EZZ'\Big)^{-1} \Big(\Delta(\theta_I) + \lambda' EXZ'\Big),$$

where $(\Delta(\theta_I), \theta_I \in \Theta_I)$ is the weak limit of the empirical process $(\Delta_n(\theta_I), \theta_I \in \Theta_I)$. For instance, under iid sampling, provided that $\{\epsilon(\theta_I)Z, \theta_I \in \Theta_I\}$ has a square-integrable envelope, the limit $\Delta(\cdot)$ is a zero mean Gaussian process with covariance kernel given by $E\Delta(\theta_I)\Delta(\theta_I')' = E\epsilon(\theta_I)\epsilon(\theta_I')ZZ'$. We therefore conclude that C.1-C.3 are satisfied and thus

$$\mathcal{C}_n = \Delta_n(\theta_I)'\Big(\frac{1}{n}\sum_{i=1}^n Z_i Z_i'\Big)^{-1} \Delta_n(\theta_I) \to_d \mathcal{C} = \sup_{\theta_I \in \Theta_I} \Delta(\theta_I)'[EZZ]^{-1}\Delta(\theta_I).$$

Notice that the limit is not pivotal and depends on knowing Θ_I.[10] Also, compactness of Θ_I is the necessary condition for the limit variable $\mathcal{C}$ to be finite.

Next, we generalize our method to the general nonlinear method of moments. Let the following partial identification condition hold: There are positive constants C and δ such that

$$\|Em_i(\theta)\|^2 \geq C \cdot \min[\inf_{\theta_I' \in \Theta_I} \|\theta - \theta_I'\|, \delta]^2, \quad \text{uniformly for } \theta \in \Theta. \tag{3.15}$$

This is both a partial identification condition and a smoothness assumption. This condition implies that $Em_i(\theta) = 0$ if an only if $\theta \in \Theta_I$, and also imposes smoothness on the behavior of $Em_i(\theta)$ for points θ near Θ_I.

In the point-identified case, global identification and the full rank and continuity of the Jacobian $\nabla_\theta Em_i(\theta)$ near θ_I ordinarily imply (3.15); see e.g. Theorem 3.3 in Pakes and Pollard (1998). In the set identified case, the Jacobian may be degenerate, which necessitates a statement of a more careful condition (3.15). For example, in the previous linear IV model we have that $Em_i(\theta) = EZ'X(\theta - \theta_I^*)$, where θ_I^* is the closest point to θ in Θ_I. Provided that $\|\theta_I^* - \theta\| > 0$, the vector $(\theta - \theta_I^*)$ is orthogonal to the hyperplane $\{v : EZ'Xv = 0\}$. Hence if rank $EZ'X$ is non-zero, for C_0 denoting the minimal positive eigenvalue of $(EX'Z)(EZ'X)$, we have $\|EZ'X(\theta - \theta_I^*)\|^2 \geq C_0 \cdot \|\theta - \theta_I^*\|^2$.

Theorem 3.4 (Generalized Method-of Moments)**.** *Let the basic conditions specified in Section 3.1 hold and let i. $\Theta_I = \partial\Theta_I$, ii. $\{m_i(\theta), \theta \in \Theta\}$ be a Donsker class, iii. $Em_i(\theta)$ satisfy (3.15) and have continuous Jacobian $G(\theta) = \nabla_\theta Em_i(\theta)$, v. $W_n(\theta) = W(\theta) + o_p(1)$ uniformly in θ, where $W(\theta)$ is*

[10] In the case $\dim(z) > \dim(x)$, and $q_n = \inf_{\theta \in \Theta} Q_n(\theta)$, then $\ell_\infty(\theta_I, \lambda) = (\Delta(\theta_I)' + \lambda' EXZ')\,[EZZ']^{-1}\,(\Delta(\theta_I) + \lambda' EXZ')$, and

$$\mathcal{C}_n \to_d \mathcal{C} = \sup_{\theta_I \in \Theta_I} \ell_\infty(\theta_I, 0) - \inf_{\theta_I \in \Theta_I, \lambda \in V_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda).$$

positive definite and continuous for all θ. Then, C.1-C.3 and A.1-A.3 hold, and

$$\ell_\infty(\theta_I, \lambda) = \left(\Delta(\theta_I) + \lambda' G(\theta_I)\right)' W(\theta_I) \left(\Delta(\theta_I) + \lambda' G(\theta_I)\right),$$

$$\mathcal{C} = \begin{cases} \sup_{\theta_I \in \Theta_I} \ell_\infty(\theta_I, 0) & \text{if } nq_n \to_p 0 \\ \sup_{\theta_I \in \Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) - \inf_{\theta_I \in \Theta_I, \lambda \in V_\infty(\theta_I)} \ \ell_\infty(\theta_I, \lambda) & \text{if } nq_n \not\to_p 0 \end{cases};$$

where $\theta_I \mapsto \Delta(\theta_I)$ is the weak limit of the process $\theta_I \mapsto \Delta_n(\theta_I)$, a zero-mean Gaussian process with the covariance function $E[\Delta(\theta_I)\Delta(\tilde{\theta}_I)'] = \lim_{n\to\infty} E[n^{-1}\sum_{i=1}^n m_i(\theta_I) \sum_{i=1}^n m_i(\tilde{\theta}_I)']$. Above, $nq_n \to_p 0$ necessarily occurs if $q_n := 0$ or if $q_n := \inf_{\theta \in \theta} Q_n(\theta)$ but $\dim(m_i(\theta)) \leq \dim(\theta)$.

The theorem verifies the conditions C.1-C.3 and A.1-A.3, which implies that the results of Theorems 2.1, 2.2, 3.1, and 3.2 apply to the GMM case. Therefore the inference procedure proposed in Section 2 is valid in this case. Further, the confidence regions have the stochastic properties given in Theorems 3.1 and 3.2. Similarly to the interval regression case, $\mathcal{C}$ depends on Θ_I. Hence the distribution of the limit variable $\mathcal{C}$ is not pivotal, and has no known closed analytical form, but this does not create a problem for the inferential method proposed in Section 2.

4. Computation and Empirical Monte Carlo

We illustrate our methods above with an empirical Monte Carlo that uses data from the CPS. We also provide details on the computational method that can be used to construct the confidence sets.

4.1. **Computation.** An issue in the subsampling method is being able to use an efficient numerical algorithm to construct level sets of an objective function. Ideally, one would like to use a rich set of grid points and evaluate the function on those points. However, as the dimension of θ increases, constructing a simple uniform grid becomes computationally infeasible. The Metropolis-Hastings algorithm provides a computationally attractive method for generating adaptive grid sets. The details of the numerical approach we use is summarized in the following algorithm:

1. Generate a grid of points $\widetilde{\Theta} = (\theta_1, ..., \theta_k)$ using the Metropolis-Hastings algorithm.[11]
2. Given a starting critical value c_0 and $k = c_0 \ln n$, we can compute the level set of the objective function as: $C_n(k) = \{\theta_g \in \widetilde{\Theta} : n(Q_n(\theta_g) - \min_{\theta_g \in C_n(k)} Q_n(\theta_g)) \leq k\}$.
3. At each subsampling stage j where $j = 1, ..., B_n$, compute $Q_b(\theta_g)$ for all $\theta_g \in C_n(k)$, and the subsample coverage statistic, e.g. $\mathcal{C}_{j,b,n} = b\big(\max_{\theta_g \in C_n(k)} Q_b(\theta_g) - \min_{\theta_g \in C_n(k)} Q_b(\theta_g)\big)$.
4. Compute the α-quantile $\hat{c}(\alpha)$ of $\{\mathcal{C}_{j,b,n}, j = 1, ..., B_n\}$.
5. Then compute $C_n(\hat{c}(\alpha)) = \{\theta_g \in \widetilde{\Theta} : n\big(Q_n(\theta_g) - \min_{\theta_g \in C_n(k)} Q_b(\theta_g)\big) \leq \hat{c}(\alpha)\}$.

[11]This algorithm is a valuable method of generating grid points adaptively, so that they are placed in relevant regions only. See Robert and Casella (1998) for a detailed description; and Chernozhukov and Hong (2003) for related examples in non-likelihood settings.

An important practical issue is the choice of the initial point c_0. Theorem 2.2 suggests the consistency of estimate c_α is not affected by the choice of c_0 as long as c_0 is bounded in the sense of (2.2). Hence a reasonable procedure for selecting c_0 can be based on pointwise testing procedures. For example, in GMM a pointwise critical value for testing $Q(\theta) = 0$ at θ is given by the α-quantile of a chi-squared variable with degrees of freedom equal to the number of equations. See the next section for how to choose c_0 in the interval regression case. More generally, one can use the α-quantile of the asymptotic distribution of the coverage statistic computed under the assumption that θ_I is a singleton.

4.2. **Empirical Monte Carlo: Returns to Schooling in the CPS.** We examine the actual finite-sample performance of the inferential procedures proposed in this paper using data from Current Population Survey. Our "population" is a sample of white men ages 20 to 50 from the March 2000 wave of the CPS. The `wages and salaries` series are not top coded or censored and so we are able to use this population to construct

1. confidence regions for the returns to schooling parameters in the point identified case, and
2. confidence regions covering the identified set when we artificially bracket the data.

Our Monte Carlo is based on random sampling from the original data set. Table 1 below provides summary statistics for our data (population). The "true" returns to schooling coefficient is .0533

Table 1. Population Summary Statistics

Variable	Obs	Mean	Std Dev	Min	Max
Wages and Salaries	13290	66667.6	51968.41	1	513472
Education	13290	11.77	1.89	1	16

with a constant term of 3.91 obtained from a least squares regression of log of wages on education in the "population". We start first by describing the details of the computational procedure.

4.3. **Starting values and Implementation Details:** The algorithm used to obtain estimates of Θ_I is the same as the one described on page 20 above. The steps of the procedure are as follows:

1. We draw a sample of size n from the above population (this population will be artificially bracketed below).
2. We build an initial estimate $C_n(c_0)$ at the starting cutoff level c_0 (choosing c_0 is described next).
3. In the subsampling step, we obtain the consistent estimate, $\hat{c}_\alpha$ of the cutoff level by subsampling the coverage statistic.

At this point one can go back to step 2 above and set $c_0 = \hat{c}_\alpha \ln n$ and then repeat step 3.[12] One may iterate several times, but we find that the cutoff levels that we get are very close after no or at most two iterations.

4. In all of the above steps we use the numerical procedures as described in Section 4.1.

The starting value c_0 that is used to construct the initial level set of the objective is problem specific. For example, in the method of moments case, the critical value from a χ^2 with an appropriate degrees of freedom can be used as the starting value. In the interval regression example, we choose an initial value c_0 is the following way. Let $Y_i^a = \frac{1}{2}(Y_{1i} + Y_{2i})$ and $\tilde{Y}_{1i} = \tilde{Y}_{2i} = Y_i^a$, then we use the objective function $Q_n(\theta)$ applied to the data $(\tilde{Y}_{1i}, \tilde{Y}_{2i}, X_i, i \leq n)$. This generates an auxiliary model, which is point-identified at some value θ_0^a, and for which we can compute the limit distribution of

$$n(Q_n(\theta_0^a) - Q_n(\hat{\theta}_0^a)),$$

and take its quantiles as the starting value c_0. In what follows, we used subsampling to estimate c_0, which is a consistent method of estimating c_0 by Theorem 2.6.1 in Politis, Romano, and Wolf (1999).

4.4. **Properties of the Set Inference Procedure in the Point-identified model.** Here, the data on wages is not interval measured (the model is point identified), but we nonetheless apply our procedure to obtain the confidence region. The objective function that we use is the minimum distance objective function

$$Q_n(\theta) = \sum_{j=1}^{J=16} \frac{n_j}{n} (\hat{\tau}_1(x_j) - x_j'\theta)_-^2 + (\hat{\tau}_2(x_j) - x_j'\theta)_+^2. \tag{4.1}$$

Since wages are not interval-measured, we have that $\hat{\tau}_1(x_j) = \hat{\tau}_2(x_j), j = 1, ..., J$. The results obtained are compared to the the joint confidence ellipse obtained using the usual Wald inference approach. In Figure 1, we provide graphs for $n = 400$ and $n = 2000$ observations and see that the subsampling procedure is close to the ellipse obtained using the usual chi-squared approximation. In Table 2, we examine the coverage of our inferential procedure . We provide the coverage for two sample sizes: $n = 1000$ and $n = 2000$, using $R = 600$ simulations. We report the coverage for a sequence of subsample sizes. As we can see, coverage seems monotonic initially in the subsample size and for the case where $n = 1000$, it peaks at $b = 200$ while for the case where $n = 2000$, it peaks at $b = 500$ and comes close to 95%.

4.5. **Properties of the Set Inference Procedure in the Set-Identified Model.** To examine the identified set of parameters in the case of censoring of the dependent variable, we bracket the income data into 15 different categories. These brackets are (in thousands)

[12] In the interval regression example and similar situations, $\ln n$ may be dropped.

FIGURE 1. Point Identified Case: Subsampling vs χ^2 Ellipses

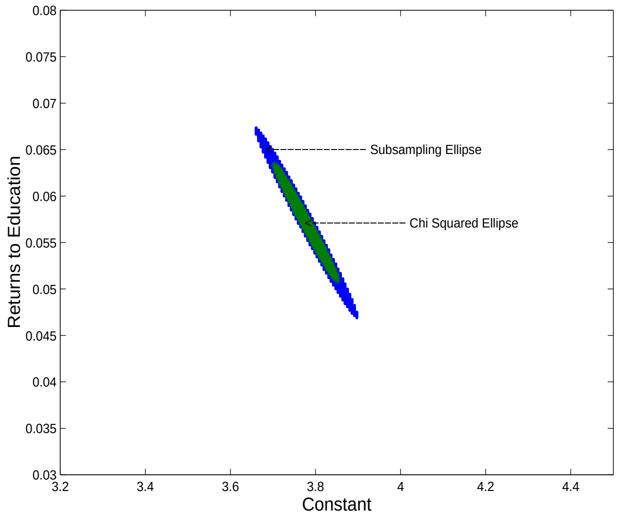


TABLE 2. **Finite-Sample Coverage Property in the Point Identified Model**

	Subsample Size				
n=1000	b=50	b=80	b=120	b=200	b=300
Coverage (95%)	85.1	88	87.2	93.1	91.2
n=2000	b=200	b=300	b=400	b=500	b=600
Coverage (95%)	86.1	88.2	90.5	95.01	93.3

```
[0,5],[5,7.5],[7.5,10],[10,12,5],[12.5,15],[15,20],[20,25],[25,30]
[30,35],[35,40],[40,50],[50,60],[60,75],[75,100],[100,150],[150,100000]
```

Notice here that the topcoding is artificially set at \$100 million. To give a flavor of the data, we obtain 718 observations in the first bracket, 1211 belong to $[50, 60]$, 1598 belong to $[100, 150]$ while 734 are making above 150. The objective function is the same as the one used earlier. In Figure 2 through our 95% confidence region $C_n(\hat{c}_{.95})$ for sample sizes $n = 600, 2000, 4000$, and 10000 drawn at random without replacement from the original "population".

Using the artificially bracketed population, we can obtain the identified set by collecting the parameters that satisfy the following set of J inequalities corresponding to the set of J values that the level of education takes:

$$E[y_1|x_j] \leq b_0 + b_1 x_j \leq E[y_2|x_j], \quad j = 1, ..., J.$$

We then graph the identified set Θ_I along with the 95% confidence region $C_n(\hat{c}_{.95})$. Notice that as the sample size increases, the set estimates shrink towards the identified set. For example, at $n = 600$, our set estimate of the intercept is $[3.2, 4.71]$ while the true range is $[3.6, 4.4]$.

FIGURE 2. Θ_I vs $C_n(\hat{c}_{.95})$: n=600, b=60.

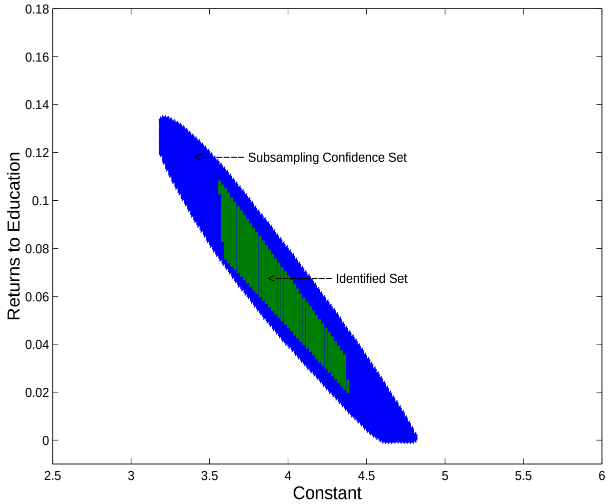


FIGURE 3. Θ_I vs $C_n(\hat{c}_{.95})$. n=2000, b=200.

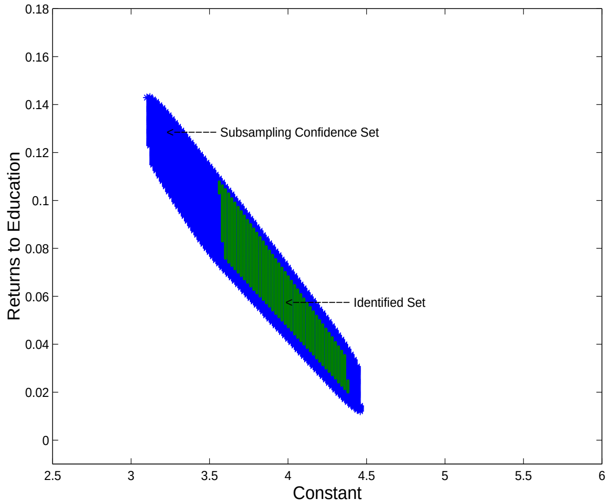


FIGURE 4. Θ_I vs $C_n(\hat{c}_{.95})$. n=4000, b=400.

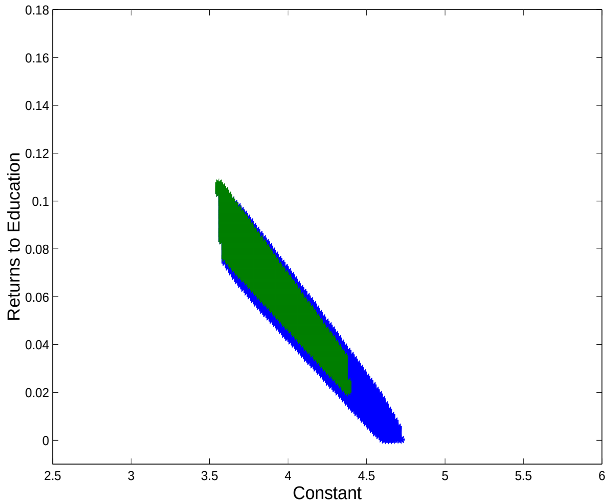
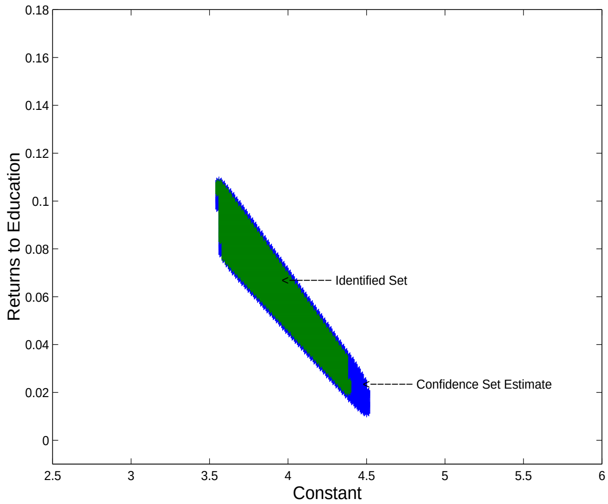


FIGURE 5. Θ_I vs $C_n(\hat{c}_{.95})$. n=10000, b=2000.



To calculate coverage probabilities, we draw $R = 600$ random samples from the population and record whether our set estimates using these samples cover the identified set. For every sample, we construct the set estimate, and calculate coverage in the following manner. Numerically, we store the

Table 3. **Finite-Sample Coverage Property in the Set Identified Model**

	Subsample Size				
N=1000	b=100	b=150	b=200	b=250	b=300
Coverage (95%)	84.1	90.2	91.3	88.5	82.8
N=2000	b=200	b=300	b=400	b=500	b=600
Coverage (95%)	86.34	90.1	92.2	94.3	91.2

identified set Θ_I and $C_n(\hat{c}_\alpha)$ as arrays, and then check whether all the points in Θ_I are contained in $C_n(\hat{c}_\alpha)$. As Table 3 and Figures 2-5 show, the coverage is similar to the point identified case and the set estimates converge to the identified set as the samples size increases. The numerical performance of our inferential methods supports the large sample theory developed in Theorems 3.1-3.3.

5. Conclusion

This paper provided confidence regions for identified sets in models with partial identification. The proposed inference procedures are criterion function based, and our confidence regions are certain level sets of the criterion function in finite samples. In the case when the model is point-identified, our confidence sets reduce to the conventional confidence regions based on inverting the likelihood or other criterion functions. The proposed procedure was shown to be valid under general yet simple conditions. Along with inferential procedures, we have developed methods of analyzing the asymptotic behavior of econometric criterion functions under set identification and also characterized the rates of convergence of the confidence regions to the identified set. We applied our methods to regressions with interval data and set identified method of moments problems. We also assessed the performance of the methods in Monte Carlo experiments based on the Current Population Survey data. We found that the methods perform well and in accordance with the asymptotic theory developed.

Appendix A. Appendix

We use the following notation for empirical processes in the sequel: for $W \equiv (Y, X)$

$$\mathbb{E}_n f(W) \equiv \frac{1}{n}\sum_{i=1}^{n} f(W_i), \quad \mathbb{G}_n f(W) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(f(W_i) - \mathbb{E} f(W_i)\right).$$

E denotes expectation and $\mathbb{E}$ denotes expectation evaluated at estimated functions $\hat{f}$:

$$\mathbb{E}\hat{f}(W_i) \equiv \left(E f(W_i)\right)_{f=\hat{f}}.$$

Outer and inner probabilities, P^* and P_* and corresponding notions of weak converegence are defined as in van der Vaart and Wellner (1996). $\to_p$ denotes convergence in outer probability, and $\to_d$ means convergence in distribution under P^*, wp $\to 1$ means "with the inner probability approaching 1," and "wp $\gtrsim 1-\epsilon$" means "with the inner probability no smaller than $1-\epsilon$ for sufficiently large n". A table of notation is given at the end of the Appendix.

Appendix B. Proofs of Lemma 2.1, Theorem 2.1, and Theorem 2.2.

B.1. **Proof of Lemma 2.1.** Proof is given in the main text. □

B.2. **Proof of Theorem 2.1.** Proof is given in the main text. □

B.3. **Proof of Theorem 2.2.** Step I. We have that

$$\widehat{\mathcal{C}}_{j,b,n} = \sup_{\theta \in C_n(\hat{c})} a_b \left(Q_{j,b}(\theta) - q_{j,b}\right), \tag{B.1}$$

where $q_{j,b} := 0$ if $q_n := 0$ and $q_{j,b} := \inf_{\theta\in\Theta} Q_{j,b}(\theta)$ otherwise; $Q_{j,b}(\theta)$ denotes the criterion function defined using the j-th subset of the data only. Define

$$\hat{G}_{b,n}(x) := B_n^{-1}\sum_{j=1}^{B_n} 1\{\widehat{\mathcal{C}}_{j,b,n} \le x\} = B_n^{-1}\sum_{j=1}^{B_n} 1\{\mathcal{C}_{j,b,n} \le x - (\widehat{\mathcal{C}}_{j,b,n} - \mathcal{C}_{j,b,n})\}, \tag{B.2}$$

where

$$\mathcal{C}_{j,b,n} = \sup_{\theta\in\Theta_I} a_b \left(Q_{j,b}(\theta) - q_{j,b}\right). \tag{B.3}$$

In what follows, the main step is to show that $\widehat{\mathcal{C}}_{j,b,n}$ can be replaced by $\mathcal{C}_{j,b,n}$. This will be possible due to the sandwich assumption, and despite that the constant $\hat{k}$ used in construction of the preliminary set estimate $C_n(\hat{k})$ is data-dependent.

By A.3 wp $\to 1$, for some deterministic set $\Theta_I^{\epsilon_n}$ we have

$$\Theta_I \subseteq C_n(\hat{k}) \subseteq \Theta_I^{\epsilon_n}, \tag{B.4}$$

where $\Theta_I^{\epsilon_n} := \{\theta_I + t : \|t\| \le \epsilon_n, \theta_I \in \Theta_I\}$. Hence wp $\to 1$ for all $i \le B_n$

$$\underbrace{\sup_{\theta\in\Theta_I} a_b \left(Q_{j,b}(\theta) - q_{j,b}\right)}_{\mathcal{C}_{j,b,n}} \le \underbrace{\sup_{\theta\in C_n(\hat{c})} a_b \left(Q_{j,b}(\theta) - q_{j,b}\right)}_{\widehat{\mathcal{C}}_{j,b,n}} \le \underbrace{\sup_{\theta\in\Theta_I^{\epsilon_n}} a_b \left(Q_{j,b}(\theta) - q_{j,b}\right)}_{\overline{\mathcal{C}}_{j,b,n}}. \tag{B.5}$$

Hence wp $\to 1$

(B.6) $$\underline{G}_{b,n}(x) \equiv B_n^{-1}\sum_{j=1}^{B_n} 1\{\overline{\mathcal{C}}_{j,b,n} \le x\} \le \widehat{G}_{b,n}(x) \le \overline{G}_{b,n}(x) \equiv B_n^{-1}\sum_{j=1}^{B_n} 1\{\mathcal{C}_{j,b,n} \le x\}.$$

By A.2 $\mathcal{C}_b \to_d \ \mathcal{C}$ and by A.3

(B.7) $$\overline{\mathcal{C}}_b = \mathcal{C}_b + o_p(1) \to_d \ \mathcal{C},$$

so that by Step II

(B.8) $$\underline{G}_{b,n}(x) \to_p \ G(x) := P\{\mathcal{C} \le x\}, \ \ \overline{G}_{b,n}(x) \to_p \ G(x) = P\{\mathcal{C} \le x\},$$

if x is a continuity point of $G(x)$, which proves that:

(B.9) $$\widehat{G}_{b,n}(x) \to_p \ G(x)$$

at the continuity points x of $G(x)$. Furthermore, if G is continuous at $c_\alpha = G^{-1}(\alpha)$,

(B.10) $$\hat{c}_\alpha = \hat{G}^{-1}_{n,b}(\alpha) \to_p \ c_\alpha = G^{-1}(\alpha),$$

since convergence of distribution function at continuity points implies the convergence of quantile functions at continuity points. Finally the claim I, that

(B.11) $$P\left(\Theta_I \in C_n(\hat{c}_\alpha)\right) \to \alpha,$$

follows by Theorem 2.1.

Step II. This part shows (B.8). Write

(B.12) $$\begin{aligned}\underline{G}_{b,n}(x) &= B_n^{-1}\sum_{j=1}^{B_n} 1\{\overline{\mathcal{C}}_{j,b,n} \le x\}\\ &\overset{(a)}{=} P\{\overline{\mathcal{C}}_b \le x\} + o_p(1)\\ &\overset{(b)}{=} P\{\mathcal{C} \ \le x\} + o_p(1)\end{aligned}$$

at the continuity points x of $G(x) \equiv P\{\mathcal{C} \ \le x\}$, as long as

(B.13) $$\frac{b}{n} \to 0, \ \ b \to \infty, \ \ n \to \infty;$$

where (a) follows from

(B.14) $$\mathrm{Var}\left(\frac{1}{B_n}\sum_{j=1}^{B_n} 1\{\overline{\mathcal{C}}_{j,b,n} \le x\}\right) = o\,(1)$$

by the variance bound for bounded U-statistics for i.i.d. series and α-mixing series given on pages 45 and 72 in Politis, Romano, and Wolf (1999); and (b) follows by

(B.15) $$\overline{\mathcal{C}}_b \to_d \ \mathcal{C}, \ \text{as } b \to \infty,$$

and the definition of convergence in distribution on $\mathbb{R}$.

Likewise, conclude

$$\text{(B.16)}\qquad \begin{aligned} \overline{G}_{b,n}(x) &= B_n^{-1}\sum_{j=1}^{B_n} 1\{\mathcal{C}_{j,b,n} \leq x\} \\ &= P\{\mathcal{C}_b \leq x\} + o_p(1) \\ &= P\{\mathcal{C} \ \leq x\} + o_p(1) \end{aligned}$$

at the continuity points x of $G(x) = P\{\mathcal{C} \ \leq x\}$.

Step III. Wp $\to 1$,

$$\text{(B.17)}\qquad \Theta_I \subset C_n(\hat{k}) \implies h(\Theta_I, C_n(\hat{k})) = 0.$$

Then, wp $\to 1$

$$\text{(B.18)}\qquad C_n(\hat{k}) \subset \Theta^{\epsilon_n} \implies h(\mathcal{C}_n(\hat{k}), \Theta_I) \leq h(\Theta^{\epsilon_n}, \Theta_I) \leq \epsilon_n.$$

Hence wp $\to 1$

$$\text{(B.19)}\qquad d_H(\Theta_I, C_n(\hat{k})) \leq \epsilon_n.$$

Thus Claim II is proven. □

Appendix C. Proof of Theorem 3.1

First recall that

$$\text{(C.1)}\qquad \begin{aligned} \mathcal{C}_n \ &:= \sup_{\theta_I \in \Theta_I} n\,(Q_n(\theta_I) - q_n) \\ &= \begin{cases} \sup_{\theta_I \in \Theta_I} \ \ell_n(\theta_I, 0) & \text{if} \quad q_n := 0, \\ \sup_{\theta_I \in \Theta_I} \ \ell_n(\theta_I, 0) - \inf_{\theta_I + \lambda/\sqrt{n} \in \Theta} \ \ell_n(\theta_I, \lambda) & \text{if} \quad q_n := \inf_{\theta \in \Theta} Q_n(\theta). \end{cases} \end{aligned}$$

We seek to establish that $\mathcal{C}_n \to_d \ \mathcal{C}$, where

$$\text{(C.2)}\qquad \mathcal{C} := \begin{cases} \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ \ell_\infty(\theta_I, \lambda), & \text{if } nq_n \to_p 0 \\ \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)}, \ \ell_\infty(\theta_I, \lambda) - \inf_{\theta_I \in \Theta_I, \lambda \in V_\infty(\theta_I)} \ \ell_\infty(\theta_I, \lambda) & \text{if } nq_n \not\to_p 0, \end{cases}$$

$$\text{(C.3)}\qquad \Lambda_\infty(\theta_I) := \{0\} \cup \{\lambda \in \mathbb{R}^d : \theta_I + \lambda/\sqrt{n} \in \text{int}(\Theta_I) \ \text{ for all sufficiently large } \ n\},$$

$$\text{(C.4)}\qquad V_\infty(\theta_I) := \{\lambda \in \mathbb{R}^d : \theta_I + \lambda/\sqrt{n} \in \Theta \text{ for all sufficiently large } n\}.$$

Step I: Case when $nq_n \to_p 0$. Since $nq_n \to_p 0$, we have that

$$\text{(C.5)}\qquad \mathcal{C}_n = \sup_{\theta_I \in \Theta_I} \ell_n\,(\theta_I, 0) + o_p(1).$$

Hence in what follows we ignore the $o_p(1)$ term.

Then we would like to show the following simple, but important representation:

$$\text{(C.6)}\qquad \mathcal{C}_n = \sup_{\theta_I \in \Theta_I} \ell_n\,(\theta_I, 0) = \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_n(\theta_I)} \ell_n(\theta_I, \lambda)$$

where

$$\text{(C.7)}\qquad \Lambda_n(\theta_I) := \{0\} \cup \{\lambda \in \mathbb{R}^d : \theta_I + \lambda/\sqrt{n'} \in \text{int}(\Theta_I) \text{ for all } n' \in [n, \infty)\}.$$

To prove (C.6), we need to show that for each $\theta \in \Theta_I$, there exists $\theta_I^* \in \partial\Theta_I$ and $\lambda^* \in \Lambda_n(\theta_I^*)$ such that $\theta_I^* + \lambda^*/\sqrt{n} = \theta_I$. If $\theta_I \in \partial\Theta_I$, then simply take $\theta_I^* = \theta_I$ and $\lambda^* = 0$. If $\theta_I \in \text{int}(\Theta_I)$ take a sufficiently small ball centered at θ_I of radius $\delta > 0$, denoted $B_\delta(\theta_I)$, such that $B_\delta(\theta_I) \subseteq \text{int}(\Theta_I)$. Then start expanding the radius of the ball δ so that to find the minimal value of δ such that the boundary of the ball includes a point $\theta_I^* \in \partial\Theta_I$. Such radius exists by compactness of Θ_I. Then the points on the line segment between θ_I^* and θ_I all belong to $\text{int}(\Theta_I)$ and can be parameterized as $\theta_I^* + \lambda^*/\sqrt{n'}$, $n' \in [n, \infty)$, where $\lambda^* = (\theta_I - \theta_I^*)\sqrt{n}$. By construction $\theta_I^* \in \partial\Theta_I$ and $\lambda^* \in \Lambda_n(\theta_I^*)$. Hence (C.6) is true.

In what follows, we distinguish two cases:

Case 1. Θ_I has an empty interior relative to Θ, so that $\Theta_I = \partial\Theta_I$,
Case 2. Θ_I has a nonempty interior relative to Θ, so that $\Theta_I \neq \partial\Theta_I$.

Case 1. By C.2-(i)

$$\mathcal{C}_n = \sup_{\theta_I \in \Theta_I} \ell_n(\theta_I, 0) = \sup_{\theta_I \in \partial\Theta_I} \ell_n(\theta_I, 0) \to_d \mathcal{C} = \sup_{\theta_I \in \partial\Theta_I} \ell_\infty(\theta_I, 0). \tag{C.8}$$

Since $\text{int}(\Theta_I)$ is empty,

$$\Lambda_n(\theta_I) = \Lambda_\infty(\theta_I) = \{0\}, \tag{C.9}$$

so that we can also rewrite (C.8) using general notation

$$\mathcal{C}_n = \sup_{\theta_I \in \Theta_I} \ell_n(\theta_I, 0) = \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_n(\theta_I)} \ell_n(\theta_I, \lambda) \to_d \mathcal{C} = \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda). \tag{C.10}$$

Case 2. For any $\delta > 0$ decompose

$$\mathcal{C}_n = \max[\mathcal{C}_n^*(\delta), \mathcal{C}_n(\delta)], \tag{C.11}$$

where

$$\mathcal{C}_n^*(\delta) := \sup_{\theta_I \in I_n(\delta)} \ell_n(\theta_I, 0) \quad \text{and} \quad \mathcal{C}_n(\delta) := \sup_{\theta_I \in \Theta_I \setminus I_n(\delta)} \ell_n(\theta_I, 0), \tag{C.12}$$

and $I_n(\delta) = \{\theta_I \in \text{int}(\Theta_I) : \inf_{\theta_I' \in \partial\Theta_I} \|\theta_I - \theta_I'\| > \delta/\sqrt{n}\}$, as defined in Assumption C.1.

By Assumption C.1 for any $\epsilon > 0$ there exists δ sufficiently large such that

$$\liminf_{n\to\infty} P_*\Big\{\mathcal{C}_n^*(\delta) = 0\Big\} \geq 1 - \epsilon. \tag{C.13}$$

Observe also that $\mathcal{C}_n(\delta) \geq 0$ by construction. Hence for any $\epsilon > 0$ there exists δ sufficiently large such that

$$\liminf_{n\to\infty} P_*\Big\{\mathcal{C}_n = \mathcal{C}_n(\delta)\Big\} \geq 1 - \epsilon. \tag{C.14}$$

Observe that

$$\mathcal{C}_n(\delta) = \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_n(\theta_I) \cap \{\|\lambda\| \leq \delta\}} \ell_n(\theta_I, \lambda). \tag{C.15}$$

By Assumption C.2 (i.-ii.) for any compact set K

$$\sup_{\theta_I \in \partial\Theta_I} \ell_n(\theta_I, \cdot) \Rightarrow \sup_{\theta_I \in \partial\Theta_I} \ell_\infty(\theta_I, \cdot) \text{ in } L^\infty(K), \tag{C.16}$$

where $\lambda \mapsto \sup_{\theta_I \in \partial\Theta_I} \ell_\infty(\theta_I, \lambda)$ has uniformly continuous paths. The next claim is that (C.16) implies by Continuous Mapping Theorem that

$$\mathcal{C}_n(\delta) \to_d \mathcal{C}(\delta) := \sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I) \cap \{\|\lambda\| \leq \delta\}} \ell_\infty(\theta_I, \lambda). \tag{C.17}$$

The proof of this claim is given below. Hence for any closed set F

$$\limsup_{n\to\infty} P^*\big\{\mathcal{C}_n(\delta) \in F\big\} \leq P\big\{\mathcal{C}(\delta) \in F\big\}. \tag{C.18}$$

Hence by (C.13) and (C.14) for any $\epsilon > 0$ there exists δ_ϵ large enough such that

$$\limsup_{n\to\infty} P^*\big\{\mathcal{C}_n \in F\big\} \leq P\big\{\mathcal{C}(\delta_\epsilon) \in F\big\} + \epsilon. \tag{C.19}$$

Therefore, taking $\epsilon \to 0$ and $\delta_\epsilon \to \infty$ accordingly, it follows that

$$\limsup_{n\to\infty} P^*\big\{\mathcal{C}_n \in F\big\} \leq P\big\{\mathcal{C} \in F\big\}, \tag{C.20}$$

where

$$\mathcal{C} := \lim_{\delta\uparrow\infty} \mathcal{C}\left(\delta\right) = \sup_{\theta_I\in\partial\Theta_I,\lambda\in\Lambda_\infty(\theta_I)} \ell_\infty\left(\theta_I,\lambda\right) \text{ a.s.} \tag{C.21}$$

The limit $\mathcal{C}$ exists in $\bar{\mathbb{R}}$ by the Monotone Convergence Theorem. By construction $\mathcal{C} \geq 0$ a.s., and by Assumption C.3(iii) $\mathcal{C} < \infty$ a.s. Conclude that by the Portmanteau lemma (van der Vaart and Wellner (1996), p.20) and (C.20) that

$$\mathcal{C}_n \to_d \mathcal{C}. \tag{C.22}$$

Proof of $\mathcal{C}_n(\delta) \to_d \mathcal{C}(\delta)$. Observe that

$$\Lambda_{n_0}(\theta_I) \subseteq \Lambda_n(\theta_I) \subseteq \Lambda_\infty(\theta_I), \text{ for all } n \geq n_0, \text{ all } n_0 \geq 1 \tag{C.23}$$

and $\Lambda_n(\theta_I) \nearrow \Lambda_\infty(\theta_I)$, in the sense that $\Lambda_n(\theta_I)$ is a nested sequence and $\Lambda_n(\theta_I) \to \Lambda_\infty(\theta_I)$, i.e.

$$\cup_{n_0=1}^\infty \cap_{n=n_0}^\infty \Lambda_n(\theta_I) = \cap_{n_0=1}^\infty \cup_{n=n_0}^\infty \Lambda_n(\theta_I) = \Lambda_\infty(\theta_I), \tag{C.24}$$

since by definitions given earlier for all $n_0 \geq 1$

$$\cup_{n=n_0}^\infty \Lambda_n(\theta_I) = \Lambda_\infty(\theta_I) \text{ and } \cap_{n=n_0}^\infty \Lambda_n(\theta_I) = \Lambda_{n_0}(\theta_I). \tag{C.25}$$

Therefore,

$$\begin{aligned} \underline{\mathcal{C}_n}(\delta) &:= \sup_{\theta_I\in\partial\Theta_I,\lambda\in\Lambda_{n_0}(\theta_I)\cap\{\|\lambda\|\leq\delta\}} \ell_n\left(\theta_I,\lambda\right) \\ &\leq \mathcal{C}_n(\delta) = \sup_{\theta_I\in\partial\Theta_I,\lambda\in\Lambda_n(\theta_I)\cap\{\|\lambda\|\leq\delta\}} \ell_n\left(\theta_I,\lambda\right) \\ &\leq \overline{\mathcal{C}_n}(\delta) := \sup_{\theta_I\in\partial\Theta_I,\lambda\in\Lambda_\infty(\theta_I)\cap\{\|\lambda\|\leq\delta\}} \ell_n\left(\theta_I,\lambda\right). \end{aligned} \tag{C.26}$$

and

$$\begin{aligned} \underline{\mathcal{C}}(\delta) &:= \sup_{\theta_I\in\partial\Theta_I,\lambda\in\Lambda_{n_0}(\theta_I)\cap\{\|\lambda\|\leq\delta\}} \ell_\infty\left(\theta_I,\lambda\right) \\ &\leq \mathcal{C}(\delta) := \sup_{\theta_I\in\partial\Theta_I,\lambda\in\Lambda_\infty(\theta_I)\cap\{\|\lambda\|\leq\delta\}} \ell_\infty\left(\theta_I,\lambda\right). \end{aligned} \tag{C.27}$$

By (C.16) and the Continuous Mapping Theorem

$$\overline{\mathcal{C}_n}(\delta) \to_d \mathcal{C}(\delta) \tag{C.28}$$

and for any fixed n_0

$$\underline{\mathcal{C}_n}(\delta) \to_d \underline{\mathcal{C}}(\delta) \tag{C.29}$$

Let the underlying probability space be denoted as $(\Omega, \mathcal{F}, P)$. Observe that for every $\omega \in \Omega$,

$$\sup_{\theta_I\in\partial\Theta_I,\lambda\in\Lambda_\infty(\theta_I)\cap\{\|\lambda\|\leq\delta\}} \ell_\infty\left(\theta_I,\lambda\right) \tag{C.30}$$

is necessarily attained at some $\theta_I^*(\omega) \in \partial\Theta_I$ and $\lambda^*(\omega) \in \Lambda_\infty(\theta_I^*(\omega)) \cap \{\|\lambda\| \leq \delta\}$, because the set $\partial\Theta_I$ is compact and $\Lambda_\infty(\theta_I^*(\omega)) \cap \{\|\lambda\| \leq \delta\}$ is also compact. That is

$$\mathcal{C}(\delta) = \ell_\infty(\theta^*(\omega), \lambda^*(\omega)) \text{ a.s.} \tag{C.31}$$

Moreover as $n_0 \to \infty$,

$$\Lambda_{n_0}(\theta_I) \cap \{\|\lambda\| \leq \delta\} \nearrow \Lambda_\infty(\theta_I) \cap \{\|\lambda\| \leq \delta\} \text{ for each } \theta_I \in \partial\Theta_I \tag{C.32}$$

hence

$$\Lambda_{n_0}(\theta_I^*(\omega)) \cap \{\|\lambda\| \leq \delta\} \nearrow \Lambda_\infty(\theta_I^*(\omega)) \cap \{\|\lambda\| \leq \delta\} \text{ a.s.} \tag{C.33}$$

so that by continuity of $\lambda \mapsto \ell_\infty(\theta_I, \lambda)$, we have that

$$\underline{\mathcal{C}}(\delta) \nearrow \lim_{n_0\to\infty} \underline{\mathcal{C}}(\delta) = \mathcal{C}(\delta) \text{ a.s.} \tag{C.34}$$

Note that to obtain (C.34) we use that $\underline{\mathcal{C}}(\delta)$ is monotonicity increasing in n_0 due to (C.32), so that $\lim_{n_0\to\infty} \underline{\mathcal{C}}(\delta)$ exists a.s. and by (C.27)

$$\lim_{n_0\to\infty} \underline{\mathcal{C}}(\delta) \leq \mathcal{C}(\delta) \text{ a.s.} \tag{C.35}$$

Moreover, for $\theta_I^*(\omega)$ defined above there exists a sequence $\lambda_{n_0}^*(\omega)$ in $\Lambda_{n_0}(\theta_I^*(\omega)) \cap \{\|\lambda\| \leq \delta\}$ such that $\lambda_{n_0}^*(\omega) \to \lambda^*(\omega)$ a.s. Hence by continuity of $\lambda \mapsto \ell_\infty(\theta_I, \lambda)$ at each $\theta_I \in \partial\Theta_I$,

$$\lim_{n_0\to\infty} \underline{\mathcal{C}}(\delta) \geq \lim_{n_0\to\infty} \ell_\infty(\theta_I^*(\omega), \lambda_{n_0}(\omega)) = \ell_\infty(\theta_I^*(\omega), \lambda^*(\omega)) = \mathcal{C}(\delta) \text{ a.s.} \tag{C.36}$$

Now we are ready to close the argument. Let F be any real number such that $P\{\mathcal{C}(\delta) = F\} = 0$, then

$$\begin{aligned} P\{\mathcal{C}(\delta) \leq F\} &\overset{(1)}{=} \lim_{n\to\infty} P^*\{\overline{\mathcal{C}}_n(\delta) \leq F\} \\ &\overset{(2)}{\leq} \liminf_{n\to\infty} P_*\{\mathcal{C}_n(\delta) \leq F\} \\ &\overset{(3)}{\leq} \limsup_{n\to\infty} P^*\{\mathcal{C}_n(\delta) \leq F\} \\ &\overset{(4)}{\leq} \limsup_{n\to\infty} P^*\{\underline{\mathcal{C}}_n(\delta) \leq F\} \\ &\overset{(5)}{\leq} P\{\underline{\mathcal{C}}(\delta) \leq F\} \text{ for any } n_0 \geq 1 \\ &\underset{n_0\to\infty}{\overset{(6)}{\to}} P\{\mathcal{C}(\delta) \leq F\}, \end{aligned} \tag{C.37}$$

where (1) is by (C.28) and the Portmanteau lemma (van der Vaart and Wellner (1996), p.20), (2)-(4) by (C.26), (5) by the Portmanteau lemma, and (6) is by (C.34) and the Portmanteau lemma. Thus for any F such that $P\{\mathcal{C}(\delta) = F\} = 0$,

$$\liminf_{n\to\infty} P_*\{\mathcal{C}_n(\delta) \leq F\} = \limsup_{n\to\infty} P^*\{\mathcal{C}_n(\delta) \leq F\} = P\{\mathcal{C}(\delta) \leq F\}. \tag{C.38}$$

Thus, by the Portmanteau lemma

$$\mathcal{C}_n(\delta) \to_d \mathcal{C}(\delta). \tag{C.39}$$

Step II.Case when $\Theta_I = \partial\Theta_I$ and $nq_n \not\to_p 0$: Consider two cases:

Case 1. Θ_I has a nonempty interior relative to Θ, so that $\Theta_I \neq \partial\Theta_I$, Case 2. Θ_I has an empty interior relative to Θ, so that $\Theta_I = \partial\Theta_I$.

Case 1. In this case by Assumption C.1

$$nq_n \to_p 0, \tag{C.40}$$

which Step I has taken care of.

Case 2. We have by definition of $\ell_n(\theta_I, \lambda)$ and $\mathcal{C}_n$,

$$\mathcal{C}_n = \sup_{\theta_I \in \Theta_I} \ell_n(\theta_I, 0) - \inf_{\theta_I + \lambda/\sqrt{n} \in \Theta} \ell_n(\theta_I, \lambda). \tag{C.41}$$

We need to show that

$$\left(\sup_{\theta_I \in \Theta_I} \ell_n(\theta_I, 0), \inf_{\theta_I + \lambda/\sqrt{n} \in \Theta} \ell_n(\theta_I, \lambda) \right) \to_d \left(\sup_{\theta_I \in \partial\Theta, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda), \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) \right). \tag{C.42}$$

Define

$$\mathcal{M}_n := \inf_{\theta_I + \lambda/\sqrt{n} \in \Theta} \ell_n(\theta_I, \lambda) \quad \text{and} \quad \mathcal{M} := \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda). \tag{C.43}$$

Below we show only the marginal convergence $\mathcal{M}_n \to_d \mathcal{M}$. The joint convergence (C.42) follows by combining the arguments of the proofs of Step I and Step II and applying the Cramer-Wold Device.

Write

$$\mathcal{M}_n = \min[\widehat{\mathcal{M}}_n(K), \mathcal{M}_n(K)], \tag{C.44}$$

where

$$\begin{aligned} \widehat{\mathcal{M}}_n(K) &:= \inf_{\theta_I + \lambda/\sqrt{n} \in \Theta,\ v(\theta_I, \lambda) \geq K/\sqrt{n}} \ell_n(\theta_I, \lambda), \\ \mathcal{M}_n(K) &:= \inf_{\theta_I + \lambda/\sqrt{n} \in \Theta,\ v(\theta_I, \lambda) \leq K/\sqrt{n}} \ell_n(\theta_I, \lambda), \end{aligned} \tag{C.45}$$

where

$$\nu(\theta_I, \lambda) := \inf_{\theta_I' \in \Theta_I} \|\theta_I + \lambda/\sqrt{n} - \theta_I'\|. \tag{C.46}$$

By Assumption C.3, for any $\epsilon > 0$, there is K large enough so that

$$\liminf_{n\to\infty} P_* \{ \widehat{\mathcal{M}}_n(K) > C_1 \min[K^2, n\delta^2] \} \geq 1 - \epsilon, \tag{C.47}$$

for some constants $C_1 > 0$ and $\delta > 0$ that do not depend on ϵ.

First observe that

$$\begin{aligned} \mathcal{M}_n(K) &= \inf_{\theta_I + \lambda/\sqrt{n} \in \Theta,\ v(\theta_I, \lambda) \leq K/\sqrt{n}} \ell_n(\theta_I, \lambda) \\ &\overset{(1)}{=} \inf_{\theta_I + \lambda/\sqrt{n} \in \Theta,\ \|\lambda\| \leq K} \ell_n(\theta_I, \lambda) \\ &\overset{(2)}{=} \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_n(\theta_I) \cap \{\|\lambda\| \leq K\}} \ell_n(\theta_I, \lambda), \end{aligned} \tag{C.48}$$

where

$$V_n(\theta_I) := \{\lambda \in \mathbb{R}^d : \theta_I + \lambda/\sqrt{n'} \in \Theta, \text{ for all } n' \in [n, \infty)\} \tag{C.49}$$

Equality (1) in (C.48) is trivial. First, by compactness of Θ_I every $\theta_I + \lambda/\sqrt{n}$ such that $\nu(\theta_I, \lambda) \leq K/\sqrt{n}$ can be represented as $\theta_I^* + \lambda^*/\sqrt{n}$ where $\lambda^* = \sqrt{n}\nu(\theta_I, \lambda) \leq K$. Second, every $\theta_I + \lambda/\sqrt{n}$ with $\|\lambda\| \leq K$ trivially satisfies $\sqrt{n}\nu(\theta_I, \lambda) \leq K$.

Equality (2) in (C.48) is more subtle. To show (2) note that $\{\theta_I + \lambda/\sqrt{n}, |\lambda| \leq K, \theta_I \in \Theta_I\} \cap \Theta = \{(\theta_I + B_{K/\sqrt{n}}(0)) \cap \Theta, \theta_I \in \Theta_I\}$. Note that $(\theta_I + B_{K/\sqrt{n}}(0)) \cap \Theta$ is *convex* and *necessarily contains* θ_I, since it is the intersection of two convex sets (by Assumption A.1 Θ is convex) that contain θ_I. Hence for every $\theta' = \theta_I + \lambda/\sqrt{n} \in (\theta_I + B_{K/\sqrt{n}}(0)) \cap \Theta$, points on the linear segment between θ' and θ_I are also contained in $(\theta_I + B_{K/\sqrt{n}}(0)) \cap \Theta$. Therefore $\lambda = \sqrt{n}(\theta' - \theta_I) \in V_n(\theta_I) \cap \{\|\lambda\| \leq K\}$, and representation (2) follows.

By Assumption C.2 (i.-ii.)

$$\inf_{\theta_I \in \Theta_I} \ell_n(\theta_I, \cdot) \Rightarrow \inf_{\theta_I \in \Theta_I} \ell_\infty(\theta_I, \cdot) \text{ in } L^\infty(K). \tag{C.50}$$

Equipped with (C.48) and (C.50), we can show through the use of the Continuous Mapping Theorem that

$$\mathcal{M}_n(K) \to_d \mathcal{M}(K) := \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_\infty(\theta_I), |\lambda| \leq K} \ell_\infty(\theta_I, \lambda). \tag{C.51}$$

The proof of this claim is stated below.

Hence by (C.47)-(C.51), for any $\epsilon > 0$ there is a sufficiently large $K(\epsilon)$ such that

$$\liminf_{n\to\infty} P_*\big\{\mathcal{M}_n = \mathcal{M}_n(K(\epsilon))\big\} \geq 1 - \epsilon. \tag{C.52}$$

Note that

$$\mathcal{M} = \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) = \lim_{K \uparrow \infty} \mathcal{M}(K). \tag{C.53}$$

$\mathcal{M}$ exists a.s. in $\bar{\mathbb{R}}$ by the Monotone Convergence Theorem. Since $\ell_\infty(\theta_I, \lambda) \geq 0$ and finite at least for $\lambda = 0$ by C.2, $0 \leq \mathcal{M} < \infty$ a.s., i.e $\mathcal{M}$ is tight.

By (C.51)-(C.52), for any $\epsilon > 0$ and each closed set F,

$$\begin{aligned} \limsup_{n\to\infty} P^*\big\{\mathcal{M}_n \in F\big\} &\leq \limsup_{n\to\infty} P^*\big\{\mathcal{M}_n(K(\epsilon)) \in F\big\} + \epsilon \\ &\leq P\big\{\mathcal{M}(K(\epsilon)) \in F\big\} + \epsilon. \end{aligned} \tag{C.54}$$

Letting $\epsilon \to 0$ and $K(\epsilon) \to \infty$ accordingly, it follows that

$$\limsup_{n\to\infty} P^*\big\{\mathcal{M}_n \in F\big\} \leq P\big\{\mathcal{M} \in F\big\}. \tag{C.55}$$

By the Portmanteau Lemma conclude that

$$\mathcal{M}_n \to_d \mathcal{M}. \tag{C.56}$$

Proof of $\mathcal{M}_n(K) \to_d \mathcal{M}(K)$. Observe that

$$V_{n_0}(\theta_I) \subseteq V_n(\theta_I) \subseteq V_\infty(\theta_I), \text{ for all } n \geq n_0 \tag{C.57}$$

so that

$$\begin{aligned} \underline{\mathcal{M}_n}(K) &:= \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_\infty(\theta_I) \cap \{\|\lambda\| \leq K\}} \ell_n(\theta_I, \lambda) \\ &\leq \mathcal{M}_n(K) = \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_n(\theta_I) \cap \{\|\lambda\| \leq K\}} \ell_n(\theta_I, \lambda) \\ &\leq \overline{\mathcal{M}_n}(K) := \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_{n_0}(\theta_I) \cap \{\|\lambda\| \leq K\}} \ell_n(\theta_I, \lambda). \end{aligned} \tag{C.58}$$

By (C.50) and the Continuous Mapping Theorem for any fixed n_0

$$\overline{\mathcal{M}_n}(K) \to_d \overline{\mathcal{M}}(K) := \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_{n_0}(\theta_I) \cap \{\|\lambda\| \leq K\}} \ell_\infty(\theta_I, \lambda), \tag{C.59}$$

and

$$\underline{\mathcal{M}}_n(K) \to_d \mathcal{M}(K) := \inf_{\theta_I \in \partial\Theta_I, \lambda \in V_\infty(\theta_I) \cap \{\|\lambda\| \le K\}} \ell_\infty(\theta_I, \lambda). \tag{C.60}$$

Moreover, as $n_0 \to \infty$

$$V_{n_0}(\theta_I) \cap \{\|\lambda\| \le K\} \nearrow V_\infty(\theta_I) \cap \{\|\lambda\| \le K\} \text{ for each } \theta_I \in \partial\Theta_I, \tag{C.61}$$

so that by continuity of $\lambda \mapsto \ell_\infty(\theta_I, \lambda)$ at each $\theta_I \in \partial\Theta_I$

$$\overline{\mathcal{M}}(K) \searrow \mathcal{M} \quad \text{a.s.} \tag{C.62}$$

The details of proving (C.62) are nearly identical to those given in (C.31)-(C.36), so are not repeated.

Let F be any real number such that $P\{\mathcal{M}(K) = F\} = 0$, then

$$\begin{aligned} P\{\mathcal{M}(K) < F\} &\overset{(1)}{=} \lim_{n\to\infty} P_*\{\underline{\mathcal{M}}_n(K) < F\} \\ &\overset{(2)}{\ge} \limsup_{n\to\infty} P^*\{\mathcal{M}_n(K) < F\} \\ &\overset{(3)}{\ge} \liminf_{n\to\infty} P_*\{\mathcal{M}_n(K) < F\} \\ &\overset{(4)}{\ge} \liminf_{n\to\infty} P_*\{\overline{\mathcal{M}}_n(K) < F\} \\ &\overset{(5)}{\ge} P\{\overline{\mathcal{M}}(K) < F\} \text{ for any } n_0 \ge 1 \\ &\underset{n_0\to\infty}{\overset{(6)}{\to}} P\{\mathcal{M}(K) < F\}, \end{aligned} \tag{C.63}$$

where (1) is by the Portmanteau lemma and (C.60), (2)-(4) by (C.58), (5) by the Portmanteau lemma, and (6) is by (C.62) and the Portmanteau lemma. Therefore, for any real F such that $P\{\mathcal{M}(K) = F\} = 0$

$$\limsup_{n\to\infty} P^*\{\mathcal{M}_n(K) < F\} = \liminf_{n\to\infty} P_*\{\mathcal{M}_n(K) < F\} = P\{\mathcal{M}(\mathcal{K}) < F\}. \tag{C.64}$$

Hence by the Portmanteau lemma

$$\mathcal{M}_n(K) \to_d \mathcal{M}(K). \tag{C.65}$$

□

Appendix D. Proof of Theorem 3.2

Steps I and II prove Claim I and Step III proves Claim II.

Step I: Case when $nq_n \to_p 0$. Recall that $Q(\theta_I) = 0$ and

$$\ell_n(\theta_I, \lambda) := n(Q_n(\theta_I + \lambda/\sqrt{n}) - Q(\theta_I)). \tag{D.1}$$

Since $nq_n \to_p 0$, we have that

$$\mathcal{C}_n = \sup_{\theta_I \in \Theta_I} \ell_n(\theta_I, 0) + o_p(1). \tag{D.2}$$

From the Proof of Theorem 3.1 we have that

$$\sup_{\theta_I \in \Theta_I} \ell_n(\theta_I, 0) = O_{p^*}(1). \tag{D.3}$$

Hence since $\hat{k} \to_p \infty$

(D.4) $$\mathcal{C}_n < \hat{k} \text{ wp} \to 1 \implies \Theta_I \subset C_n(\hat{k}) \text{ wp} \to 1 \implies h(\Theta_I, C_n(\hat{k})) = 0 \text{ wp} \to 1.$$

Condition C.3 implies that for any $K \to \infty$ with $K/\hat{k} \to 0$ there are positive constants C_1 and δ such that wp $\to 1$

(D.5) $$C_1 \cdot n \cdot \min[\nu(\theta_I, \lambda)^2, \delta^2] \leq \ell_n(\theta_I, \lambda)$$

uniformly in (θ_I, λ) such that $\nu(\theta_I, \lambda) \geq K/\sqrt{n}$, where $\nu(\theta_I, \lambda) = \inf_{\theta_I' \in \Theta_I} \|\theta_I + \lambda/\sqrt{n} - \theta_I'\|$.

By definition of $C_n(\hat{k})$ and since $Q(\theta_I) = 0$

(D.6) $$\sup_{\theta \in C_n(\hat{k})} nQ_n(\theta) + o_p(1) = \sup_{\theta_I + \lambda/\sqrt{n} \in C_n(\hat{k})} \ell_n(\theta_I, \lambda) + o_p(1) \leq \hat{k}.$$

Hence

(D.7) $$\theta_I + \lambda/\sqrt{n} \in C_n(\hat{k}) \text{ implies } \ell_n(\theta_I, \lambda) \leq \hat{k} + o_p(1).$$

Then the claim is that wp $\to 1$

(D.8) $$C_n(\hat{k}) \subset \Theta_I^{2(\hat{k}/C_1)^{1/2}/\sqrt{n}},$$

where $\Theta_I^c := \{\theta_I + t : \|t\| \leq c, \theta_I \in \Theta_I\}$. Suppose otherwise, then for some $\theta_I + \lambda/\sqrt{n} \in C_n(\hat{k})$,

(D.9) $$\nu(\theta_I, \lambda) = \inf_{\theta_I' \in \Theta_I} |\theta_I + \lambda/\sqrt{n} - \theta_I'| > 2(\hat{k}/C_1)^{1/2}/\sqrt{n}.$$

Then for this pair (θ_I, λ), wp $\to 1$

(D.10) $$4\hat{k} \overset{(a)}{<} C_1 \cdot n \cdot \min[\nu(\theta_I, \lambda)^2, \delta^2] \overset{(b)}{\leq} \ell_n(\theta_I, \lambda) \overset{(c)}{\leq} \hat{k} + o_p(1),$$

where (a) is by (D.9) and by $\hat{k} \to \infty$ and $\hat{k}/n \to 0$ so that $C_1 \cdot n \cdot \min[\nu(\theta_I, \lambda)^2, \delta^2] \geq C_1 \cdot n \cdot \min[4(\hat{k}/C_1)/n, \delta^2] \geq 4\hat{k}$ wp $\to 1$, (b) is by (D.5), (c) is by (D.7). This yields a contradiction. Thus, (D.8) is true.

Combining (D.4) and (D.8) it follows that wp $\to 1$

(D.11) $$d_H(\Theta_I, C_n(\hat{k})) \leq h(C_n(\hat{k}), \Theta_I) \leq h(\Theta_I^{2(\hat{k}/C_1)^{1/2}/\sqrt{n}}, \Theta_I) \leq 2(\hat{k}/C_1)^{1/2}/\sqrt{n}.$$

Step II: Case when $nq_n \not\to_p 0$. Observe the relationship between the "centered" and "un-centered" confidence regions. Let the un-centered region be denoted as before:

(D.12) $$C_n(c) := \Big\{\theta : n\left(Q_n(\theta)\right) \leq c\Big\},$$

and the centered version be denoted as (in this proof only):

(D.13) $$\widetilde{C}_n(c) := \Big\{\theta : n\left(Q_n(\theta) - q_n\right) \leq c\Big\}, \text{ where } q_n := \inf_{\theta \in \Theta} Q_n(\theta).$$

Observe that the following key relationship between the two sets

(D.14) $$\widetilde{C}_n(k) = C_n(k + nq_n) \text{ for all } k > 0.$$

Let $\hat{k} \to \infty$ but $\hat{k} = o_p(n)$. From the proof of Theorem 2.1 we know that

(D.15) $$\mathcal{M}_n = nq_n = O_p(1) \text{ since } \mathcal{M}_n \to_d \mathcal{M}.$$

Hence

(D.16) $$\widetilde{k} := (\hat{k} + nq_n) = \hat{k} \cdot (1 + o_p(1)).$$

By (D.14)

$$\widetilde{C}_n(\hat{k}) = C_n(\widetilde{k}), \quad \text{so} \quad d_H(\widetilde{C}_n(\hat{k}), \Theta_I) = d_H(C_n(\widetilde{k}), \Theta_I). \tag{D.17}$$

Hence by (D.11) and (D.16) it follows that wp $\to 1$

$$d_H(\widetilde{C}_n(\hat{k}), \Theta_I) \le d_H(C_n(\widetilde{k}), \Theta_I) \le 2(\tilde{k}/C_1)^{1/2}/\sqrt{n} \le \left(2(\hat{k}/C_1)^{1/2}/\sqrt{n}\right)(1 + o_p(1)), \tag{D.18}$$

and the claim now follows.

Step III proves Claim II, that for $\hat{k} \in [c_0, c_1] \cdot \ln n$, wp $\to 1$

$$\Theta_I \subset C_n(\hat{k}) \subset \Theta_I^{\epsilon_n}, \text{ where } \; b(\sup_{\theta \in \Theta_I^{\epsilon_n}} Q_b(\theta) - \sup_{\theta \in \Theta_I} Q_b(\theta)) = o_p(1), \tag{D.19}$$

$\Theta_I^{\epsilon_n} = \{\theta_I + t : \|t\| \le \epsilon_n, \theta_I \in \Theta_I\}$, and $\epsilon_n \to 0$ is a sequence of positive constants.

We have that $\hat{k} \in [c_0, c_1] \cdot \ln n$ wp $\to 1$. Hence, by Claim I

$$\Theta_I \subset C_n(\hat{k}) \subset \Theta_I^{\epsilon_n} \quad \text{wp } \to 1, \text{ where } \epsilon_n = C \cdot \ln n/\sqrt{n} \tag{D.20}$$

for some $C > 0$. Then, using $Q(\theta_I) = 0$,

$$\begin{aligned}
0 \le b(\sup_{\theta \in \Theta_I^{\epsilon_n}} Q_b(\theta) - \sup_{\theta_I \in \Theta_I} Q_b(\theta_I)) & \\
&= \sup_{\theta_I + \lambda/\sqrt{b} \in \Theta_I^{\epsilon_n}} b(Q_b(\theta_I + \lambda/\sqrt{b} - Q(\theta_I)) - \sup_{\theta_I \in \Theta_I} b(Q_b(\theta_I)) - Q(\theta_I)) \\
&\le \sup_{\theta_I \in \Theta_I, |\lambda| \le \sqrt{b}\epsilon_n} \ell_b(\theta_I, \lambda) - \sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0) \\
&= \max[\sup_{\theta_I \in \partial\Theta_I, |\lambda| \le \sqrt{b}\epsilon_n} \ell_b(\theta_I, \lambda), \sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0)] - \sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0) \\
&\overset{(a)}{=} \max[\sup_{\theta_I \in \partial\Theta_I} \ell_b(\theta_I, 0) + o_p(1), \sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0)] - \sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0) \\
&\overset{(b)}{=} \max[\sup_{\theta_I \in \partial\Theta_I} \ell_b(\theta_I, 0), \sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0)] + o_p(1) - \sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0) \\
&= \sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0) + o_p(1) - \sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0) \\
&= o_p(1),
\end{aligned} \tag{D.21}$$

where (a) follows from the uniform stochastic equicontinuity of the process $\lambda \mapsto \sup_{\theta_I \in \partial\Theta_I} \ell_b(\theta_I, \lambda)$ over K assumed in C.2 and from

$$\sqrt{b}\epsilon_n \to 0, \tag{D.22}$$

which follows from the assumption that

$$b/n \to 0 \quad \text{at polynomial rate, so that} \quad \sqrt{b}\epsilon_n \propto \sqrt{b/n}\ln n \to 0; \tag{D.23}$$

(b) follows by the Continuous Mapping Theorem and proof of Theorem 3.1 which implies that as $b \to \infty$

$$\left(\sup_{\theta_I \in \Theta_I} \ell_b(\theta_I, 0), \sup_{\theta_I \in \partial\Theta_I} \ell_b(\theta_I, 0)\right) \to_d \left(\sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda), \sup_{\theta_I \in \partial\Theta_I} \ell_\infty(\theta_I, 0)\right). \tag{D.24}$$

□

Appendix E. Proof of Theorem 3.3

Verification of A.1 is immediate from the stated assumptions. Therefore we will focus on verifying C.1 to C.3.

Step I. Verification of C.1: The identified set Θ_I is determined by a set of inequalities: $\Theta_I := \{\theta \in \mathbb{R}^d : x_j'\theta \geq \tau_1(x_j)$ and $x_j'\theta \leq \tau_2(x_j)$ for all $j \leq J\}$. It is assumed that the $d \times J$ matrix $\mathcal{X} := (x_j, j \leq J)$ is of rank $d = \dim(\theta)$. Denote $\mathcal{T}_k := (\tau_k(x_j), j \leq J)$ for $k = 1, 2$. Thus

$$\Theta_I = \{\theta \in \mathbb{R}^d : \mathcal{X}'\theta = t \text{ for some } t : \mathcal{T}_1 \leq t \leq \mathcal{T}_2\}. \tag{E.1}$$

Since $\mathcal{T} = \{t \in \mathbb{R}^J : \mathcal{T}_1 \leq t \leq \mathcal{T}_2\}$ is compact and $\mathcal{X}$ has full rank, Θ_I is compact. It is assumed that $\Theta_I \subset \Theta$. $\partial\Theta_I$ is determined as

$$\partial\Theta_I = \{\theta \in \Theta_I : x_j'\theta = \tau_1(x_j) \text{ or } x_j'\theta = \tau_2(x_j), \text{ for some } j\}. \tag{E.2}$$

That $\theta_I \in I_n(\delta) = \{\theta_I \in \mathrm{int}(\Theta_I) : \inf_{\theta_I' \in \partial\Theta_I} \|\theta_I - \theta_I'\| \geq \delta/\sqrt{n}\}$ implies that $x_j'\theta_I$ must be bounded away from any $\tau_k(x_j)$ by the distance proportional to $\|x_j\|\delta/\sqrt{n}$. Indeed, observe that there exists $\kappa > 0$ such that

$$\frac{\|x_j'\theta_I' - x_j'\theta_I\|}{\|x_j\|\|\theta_I' - \theta_I\|} \geq \kappa > 0 \quad \forall\theta_I \notin \partial\Theta_I, \forall\theta_I' \in \partial\Theta_I : x_j'\theta_I' = \tau_k(x_j), \quad \forall(j,k). \tag{E.3}$$

Suppose otherwise that $\kappa = 0$. This implies $x_j'\theta_I = x_j'\theta_I' = \tau_k(x_j)$ for some j and k and some $\theta_I' \in \partial\Theta_I$. This poses a contradiction to $\theta_I \notin \partial\Theta_I$. Hence $\kappa > 0$. Hence

$$\frac{\|\tau_k(x_j) - x_j'\theta_I\|}{\|x_j\|} \geq \kappa\|\theta_I' - \theta_I\| \quad \forall\theta_I \notin \partial\Theta_I, \forall\theta_I' \in \partial\Theta_I : x_j'\theta_I' = \tau_k(x_j), \quad \forall(j,k). \tag{E.4}$$

Recall that $\|x_j\| \geq 1$ for all j since the first component of x_j is 1. Hence

$$\|\tau_k(x_j) - x_j'\theta_I\| \geq \kappa \inf_{\theta_I' \in \partial\Theta_I} \|\theta_I' - \theta_I\|\|x_j\| \quad \forall\theta_I \notin \partial\Theta_I, \quad \forall(j,k). \tag{E.5}$$

Thus,

$$\theta_I \in I_n(\delta) \quad \Rightarrow \quad \tau_1(x_j) - x_j'\theta_I \leq -\frac{\kappa\delta}{\sqrt{n}}\|x_j\| \text{ or } \tau_2(x_j) - x_j'\theta_I \geq \frac{\kappa\delta}{\sqrt{n}}\|x_j\|, \quad \forall j. \tag{E.6}$$

Recall

$$\ell_n(\theta_I, \lambda) = \sum_{j=1}^{J} \frac{n_j}{n}\left[\sqrt{n}\left(\hat{\tau}_1(x_j) - x_j'\theta_I\right) - x_j'\lambda\right]_+^2 + \sum_{j=1}^{J} \frac{n_j}{n}\left[\sqrt{n}\left(\hat{\tau}_2(x_j) - x_j'\theta_I\right) - x_j'\lambda\right]_-^2. \tag{E.7}$$

Hence

$$\ell_n(\theta_I, 0) = \sum_{j=1}^{J} \frac{n_j}{n}\left[\hat{W}_{1j} - \sqrt{n}\left(x_j'\theta_I - \tau_1(x_j)\right)\right]_+^2 + \sum_{j=1}^{J} \frac{n_j}{n}\left[\hat{W}_{2j} - \sqrt{n}\left(x_j'\theta_I - \tau_2(x_j)\right)\right]_-^2. \tag{E.8}$$

Hence

$$\sup_{\theta_I \in I_n(\delta)} |\ell_n(\theta_I, 0)| \leq \sum_{j=1}^{J} \frac{n_j}{n}\left[\hat{W}_{1j} - \kappa\delta\|x_j\|\right]_+^2 + \sum_{j=1}^{J} \frac{n_j}{n}\left[\hat{W}_{2j} + \kappa\delta\|x_j\|\right]_-^2. \tag{E.9}$$

Choose any $\epsilon > 0$. Since $\|x_j\| \geq 1$, $W_{1j} = O_p(1)$ and $W_{2j} = O_p(1)$ for each $j \leq J$, for δ sufficiently large,

$$\limsup_{n\to\infty} \max_{j \leq J} P(W_{1j} \geq \kappa\delta\|x_j\|) \leq \epsilon \quad \text{and} \quad \max_{j \leq J} \limsup_{n\to\infty} P(W_{2j} \leq -\kappa\delta\|x_j\|) \leq \epsilon. \tag{E.10}$$

C.1 is now verified, since it follows that for any $\epsilon > 0$, there exists δ sufficiently large such that

$$\sup_{\theta_I \in I_n(\delta)} |\ell_n(\theta_I, 0)| = 0 \text{ wp } \gtrsim 1 - \epsilon. \tag{E.11}$$

Step II. Verification of part i of C.2: Write $\ell_n(\theta_I, \lambda) = \ell_n^{(1)}(\theta_I, \lambda) + \ell_n^{(2)}(\theta_I, \lambda)$, where

$$\begin{aligned} \ell_n^{(1)}(\theta_I, \lambda) &= \sum_{j=1}^{J} \frac{n_j}{n} \left[\hat{W}_{1j} - x_j'\lambda\right]_+^2 1\left(x_j'\theta_I = \tau_1(x_j)\right) \\ &\quad + \sum_{j=1}^{J} \frac{n_j}{n} \left[\hat{W}_{2j} - x_j'\lambda\right]_-^2 1\left(x_j'\theta_I = \tau_1(x_j)\right) \\ \ell_n^{(2)}(\theta_I, \lambda) &= \sum_{j=1}^{J} \frac{n_j}{n} \left[\hat{W}_{1j} - x_j'\lambda - \sqrt{n}\left(x_j'\theta_I - \tau_1(x_j)\right)\right]_+^2 1\left(x_j'\theta_I > \tau_1(x_j)\right) \\ &\quad + \sum_{j=1}^{J} \frac{n_j}{n} \left[\hat{W}_{2j} - x_j'\lambda - \sqrt{n}\left(x_j'\theta_I - \tau_2(x_j)\right)\right]_-^2 1\left(x_j'\theta_I < \tau_2(x_j)\right). \end{aligned} \tag{E.12}$$

Note first that for fixed $\theta_I \in \partial\Theta_I$ and fixed λ, by an argument similar to that in Step I,

$$\ell_n^{(2)}(\theta_I, \lambda) = 0 \text{ wp } \to 1. \tag{E.13}$$

Therefore, by the Continuous Mapping Theorem the finite-dimensional limit of $\ell_n(\theta_I, \lambda)$ is given by

$$\ell_\infty(\theta_I, \lambda) \equiv \sum_{j=1}^{J} p_j \left[W_{1j} - x_j'\lambda\right]_+^2 1\left(x_j'\theta_I = \tau_1(x_j)\right) + \sum_{j=1}^{J} p_j \left[W_{2j} - x_j'\lambda\right]_-^2 1\left(x_j'\theta_I = \tau_2(x_j)\right). \tag{E.14}$$

Step III. Verification of part ii(a) of C.2: First, suppose that $\Theta_I \neq \partial\Theta_I$, that is the interior of Θ_I is nonempty relative to $\mathbb{R}^d$. C.2-ii(a) requires us to examine the finite-dimensional limit theory of $\sup_{\theta_I \in \partial\Theta_I} \ell_n(\theta_I, \cdot)$. The problem of finding $\sup_{\theta_I \in \partial\Theta_I} \ell_n(\theta_I, \lambda)$ for $n \leq \infty$ amounts to choosing θ_I among the finite subset of points $V_I \subset \Theta_I$ that are defined as a collection of solutions to all systems of d-equations of the form:

$$x_{j_l}'\theta_I = \tau_{k_l}(x_{j_l}), \quad l = 1, ..., d, \tag{E.15}$$

such that $(x_{j_l}, l = 1, ..., d)$ has rank d and $(k_l, j_l) \in \{1, 2\} \times \{1, 2, ..., J\}$ for each l. There is only a finite set of such systems of equations. Then,

$$\sup_{\theta_I \in \partial\Theta_I} \ell_n(\theta_I, \lambda) = \max_{\theta_I \in V_I} \ell_n(\theta_I, \lambda), \text{ for } n \leq \infty. \tag{E.16}$$

Hence by the Continuous Mapping Theorem the finite-dimensional weak limit of $\max_{\theta_I \in V_I} \ell_n(\theta_I, \cdot)$ is given by $\max_{\theta_I \in V_I} \ell_\infty(\theta_I, \cdot)$, and therefore the finite-dimensional weak limit of $\sup_{\theta_I \in \partial\Theta_I} \ell_n(\theta_I, \cdot)$ is given by $\sup_{\theta_I \in \partial\Theta_I} \ell_\infty(\theta_I, \cdot)$.

Second, suppose that $\Theta_I = \partial\Theta_I$, that is the interior of Θ_I empty relative to $\mathbb{R}^d$. Then wp $\to 1$

$$\begin{aligned} \inf_{\theta_I \in \Theta_I} \ell_n(\theta_I, \lambda) &= \sum_{j \leq J} 1\left(\tau_1(x_j) = \tau_2(x_j)\right) \frac{n_j}{n} \left(\hat{W}_{1j} - x_j'\lambda\right)^2 \\ \inf_{\theta_I \in \Theta_I} \ell_\infty(\theta_I, \lambda) &= \sum_{j \leq J} 1\left(\tau_1(x_j) = \tau_2(x_j)\right) p_j \left(W_{1j} - x_j'\lambda\right)^2. \end{aligned} \tag{E.17}$$

Therefore, by Continuous Mapping Theorem the finite-dimensional limit of $\inf_{\theta_I \in \Theta_I} \ell_n(\theta_I, \cdot)$ is given by $\inf_{\theta_I \in \Theta_I} \ell_\infty(\theta_I, \cdot)$. The joint finite-dimensional convergence of $\left(\inf_{\theta_I \in \Theta_I} \ell_n(\theta_I, \cdot), \sup_{\theta_I \in \Theta_I} \ell_n(\theta_I, \cdot)\right)$ follows by the Continuous Mapping Theorem.

Step III. Verification of part ii(b) of C.2: Let K be any compact subset of $\mathbb{R}^d$. To verify the stochastic equicontinuity for $\max_{\theta_I \in V_I} \ell_n(\theta_I, \lambda)$ over $\lambda \in K$, where V_I is a *finite* subset of $\partial\Theta_I$, it suffices to show that for any fixed $\theta_I \in \partial\Theta_I$, $\ell_n(\theta_I, \lambda)$ is stochastically equicontinuous in $\lambda \in K$. This immediately follows from convexity of $\ell_n(\theta_I, \cdot)$ in λ and finite-dimensional convergence of $\ell_n(\theta_I, \cdot)$ to a convex continuous function $\ell_\infty(\theta_I, \cdot)$ over $\mathbb{R}^d$, proven in Step I. Likewise, stochastic equicontinuity of $\inf_{\theta_I \in \Theta_I} \ell_n(\theta_I, \lambda)$ in $\lambda \in K$ also follows from convexity and finite-dimensional convergence of $\inf_{\theta_I \in \Theta_I} \ell_n(\theta_I, \cdot)$ to a convex continuous function $\inf_{\theta_I \in \Theta_I} \ell_\infty(\theta_I, \cdot)$, cf. (E.17).

Step IV. Verification of part iii of C.2: Any $\lambda \in \Lambda_\infty(\theta_I)$ for $\theta_I \in \partial\Theta_I$ satisfies the property that $x_j'\lambda \geq 0$ whenever $\tau_1(x_j) = x_j'\theta_I$, and $x_j'\lambda \leq 0$ whenever $\tau_2(x_j) = x_j'\theta_I$. Hence

$$\sup_{\theta_I \in \partial\Theta_I, \lambda \in \Lambda_\infty(\theta_I)} \ell_\infty(\theta_I, \lambda) \leq \sum_{j=1}^{J} p_j \left([W_{1j}]_+^2 + [W_{2j}]_-^2 \right) < \infty \text{ a.s.} \tag{E.18}$$

Step V. Verification of C.3: For each $\theta_n(\theta_I, \lambda)$ such that $\nu(\theta_I, \lambda) = \inf_{\theta_I' \in \Theta_I} \|\theta_I + \lambda/\sqrt{n} - \theta_I'\| \geq K/\sqrt{n}$, decompose

$$\theta_I + \lambda/\sqrt{n} = \theta_I^* + \lambda^*/\sqrt{n}, \text{ where } \theta_I^* = \arg \inf_{\theta \in \partial\Theta_I} \|\theta_I + \lambda/\sqrt{n} - \theta\| \text{ and } \|\lambda^*\| = \sqrt{n}v(\theta_I, \lambda) \geq K. \tag{E.19}$$

Any solution θ_I^* is subject to $1 \leq p \leq \dim(\theta)$ binding constraints of the form:

$$x_{j_l}'\theta_I^* = \tau_{k_l}(x_{j_l}), \quad l = 1, ..., p, \tag{E.20}$$

where $\mathcal{X}^* := (x_{j_l}, l = 1, ..., p)$ has rank p and $(k_l, j_l) \in \{1, 2\} \times \{1, ..., J\}$ for each l. Matrix $\mathcal{X}^{*\prime}\mathcal{X}^*$ (which depends on (θ_I, λ)) is necessarily one of the finitely many $p \times p$ sub-matrices of $\mathcal{X}^{*\prime}\mathcal{X}^*$ that have full rank p, and whose eigenvalues are bounded above away from zero. Let also

$$\mathcal{T}^* = (\tau_{k_l}(x_{j_l}), l = 1, ..., p) \text{ and } \mathcal{JK}^* = ((j_l, k_l), l = 1, ..., p). \tag{E.21}$$

Since the eigenvalues of $\left(\mathcal{X}^{*\prime}\mathcal{X}^*\right)^{-1}$ are bounded away from zero,

$$c_1\|\lambda^*\| \leq \|\mathcal{X}^{*\prime}\lambda^*\|_\infty. \tag{E.22}$$

Moreover, given any index set $\mathcal{JK}^*$, for *all* $(j_l, k_l) \in \mathcal{JK}^*$:

$$x_{j_l}'\lambda^* \leq 0 \quad \text{if} \quad k_l = 1 \quad \text{or} \quad x_{j_l}'\lambda^* \geq 0 \quad \text{if} \quad k_l = 2. \tag{E.23}$$

By (E.22) *at least for one* $(j^*, k^*) \in \mathcal{JK}^*$:

$$c_1\|\lambda^*\| \leq |x_{j^*}'\lambda^*| \quad \text{so that } x_{j^*}'\lambda^* \leq -c_1\|\lambda^*\| \quad \text{if} \quad k^* = 1 \quad \text{or} \quad x_{j^*}'\lambda^* \geq c_1\|\lambda^*\| \quad \text{if} \quad k^* = 2. \tag{E.24}$$

Decompose $\ell_n(\theta_I,\lambda) = \ell_n^{(1)}(\theta_I,\lambda) + \ell_n^{(2)}(\theta_I,\lambda)$, where

(E.25)

$$\begin{aligned}
\ell_n^{(1)}(\theta_I,\lambda) =& \sum_{j=1}^{J} \frac{n_j}{n}\left[\hat{W}_{1j} - \sqrt{n}\left(x_j'\lambda^*\right)\right]_+^2 1\left(x_j'\theta_I^* = \tau_1(x_j)\right) + \frac{n_j}{n}\left[\hat{W}_{2j} - \sqrt{n}\left(x_j'\lambda^*\right)\right]_-^2 1\left(x_j'\theta_I^* = \tau_2(x_j)\right) \\
\geq& \frac{n_{j^*}}{n}\left[\hat{W}_{1j^*} - \sqrt{n}\left(x_{j^*}'\lambda^*\right)\right]_+^2 1(k^*=1) + \left[\sqrt{n}\left(x_{j^*}'\lambda^*\right) - \hat{W}_{2j^*}\right]_+^2 1(k^*=2) \\
\geq& p_{j^*}(1+o_p(1))\left[\hat{W}_{1j^*} + c_1\|\lambda^*\|\right]_+^2 1(k^*=1) + \left[c_1\|\lambda^*\| - \hat{W}_{2j^*}\right]_+^2 1(k^*=2) \\
\geq& p_{j^*}(1+o_p(1))\left[\min[\hat{W}_{1j^*}, -\hat{W}_{2j^*}] + c_1\|\lambda^*\|\right]_+^2 \\
\geq& \min_{j\leq J} p_j(1+o_p(1))\left[\min[\hat{W}_{1j}, -\hat{W}_{2j}, j\leq J] + c_1\|\lambda^*\|\right]_+^2,
\end{aligned}$$

and

(E.26)

$$\begin{aligned}
\ell_n^2(\theta_I,\lambda) =& \sum_{j=1}^{J}\frac{n_j}{n}\left[\hat{W}_{1j} - \sqrt{n}(x_j'\theta^* - \tau_1(x_j)) - x_j'\lambda^*\right]_+^2 1(x_j'\theta_I^* > \tau_1(x_j)) \\
&+ \sum_{j=1}^{J}\frac{n_j}{n}\left[\hat{W}_{2j} - \sqrt{n}(x_j'\theta^* - \tau_2(x_j)) - x_j'\lambda^*\right]_-^2 1(x_j'\theta_I^* < \tau_2(x_j)) \geq 0.
\end{aligned}$$

Hence, recalling that $\|\lambda^*\| = \sqrt{n}\nu(\theta_I,\lambda)$, one has that $\ell_n^{(1)}(\theta_I,\lambda) \geq c\,(\underline{W} + c'\|\lambda^*\|)_+^2$, for some positive constants $c>0$ and $c'>0$, where $\underline{W} := \min\left[\hat{W}_{1j}, -\hat{W}_{2j}, j\leq J\right]$. Let $\sqrt{n}v(\theta_I,\lambda) = \|\lambda^*\| \geq K$. Since $|\underline{W}| = O_p(1)$, for any $\epsilon>0$, we can select K large enough so that $\underline{W} > -c'\sqrt{n}v(\theta_I,\lambda)/2$ wp $\gtrsim 1-\epsilon$. Then wp $\gtrsim 1-\epsilon$,

(E.27)
$$\ell_n^{(1)}(\theta_I,\lambda) \geq \frac{1}{2}cc'v(\theta_I,\lambda)^2.$$

Since $\ell_n(\theta_I,\lambda) \geq \ell_n^{(1)}(\theta_I,\lambda)$, C.3 is verified. □

Appendix F. Proof of Theorem 3.4

Step I. Verification of A.1 is immediate from the stated assumptions and from $\{m_i(\theta), \theta\in\Theta\}$ being Donsker (hence Glivenko-Cantelli).

Step II. Verification of C.1 is not needed because $\Theta_I = \partial\Theta_I$ in Θ.

Step III. Verification of C.2 Write

(F.1)

$$\begin{aligned}
\ell_n(\theta_I,\lambda) \equiv& \sqrt{n}\mathbb{E}_n m_i(\theta_I+\lambda/\sqrt{n})' W_n(\theta_I+\lambda/\sqrt{n})\sqrt{n}\mathbb{E}_n m_i(\theta_I+\lambda/\sqrt{n}) \\
\equiv& \left(\mathbb{G}_n m_i(\theta_I+\lambda/\sqrt{n}) + \sqrt{n}E m_i(\theta_I+\lambda/\sqrt{n})\right)' \\
&\times W_n(\theta_I+\lambda/\sqrt{n})\left(\mathbb{G}_n m_i(\theta_I+\lambda/\sqrt{n}) + \sqrt{n}E m_i(\theta_I+\lambda/\sqrt{n})\right).
\end{aligned}$$

By condition **ii.** that $\{m_i(\theta), \theta\in\Theta\}$ forms a Donsker class for any compact set K,

(F.2)
$$\mathbb{G}_n m_i(\theta_I+\lambda/\sqrt{n}) = \mathbb{G}_n m_i(\theta_I) + o_p(1) \Rightarrow \Delta(\theta_I), \text{ in } L^\infty(\Theta_I\times K)$$

where $\Delta(\theta_I)$ is the Gaussian process with the covariance functions given in the statement of Theorem 2.4. By condition **iv.**,

(F.3)
$$W_n(\theta_I+\lambda/\sqrt{n}) = W(\theta_I) + o_p(1), \text{ in } L^\infty(\Theta_I\times K).$$

By condition **iii.**,

$$\sqrt{n}Em_i(\theta_I + \lambda/\sqrt{n}) = G(\theta_I)'\lambda + o(1), \text{ in } L^\infty(\Theta_I \times K). \tag{F.4}$$

Hence

$$\ell_n(\theta_I, \lambda) \Rightarrow \ell_\infty(\theta_I, \lambda) \equiv (\Delta(\theta_I) + G(\theta_I)'\lambda)' \times W(\theta_I) \times (\Delta(\theta_I) + G(\theta_I)'\lambda), \text{ in } L^\infty(\Theta_I \times K). \tag{F.5}$$

Hence C.2-i. is verified, since finite-dimensional convergence is implied by weak convergence in $L^\infty(\Theta_I \times K)$. C.2-ii.(a) is verified by the Continuous Mapping Theorem since the functionals $l_n(\lambda) = \inf_{\theta_I \in \Theta_I} \ell_n(\theta_I, \lambda)$ and $u_n(\lambda) = \sup_{\theta_I \in \Theta_I} \ell_n(\theta_I, \lambda)$ are continuous transformations of $\ell_n(\theta, \lambda)$. C.2-ii.(b) follows as well, since stochastic equicontinuity of $l_n(\lambda)$ and $u_n(\lambda)$ is implied by stochastic equicontinuity of $\ell_n(\theta_I, \lambda)$ over $\Theta_I \times K$, which is implied by weak convergence of $\ell_n(\theta_I, \lambda)$ in $L^\infty(\Theta_I \times K)$ to the quadratic form of a Gaussian process, $\ell_\infty(\theta_I, \lambda)$.

Step IV. Verification of C.3. By condition **ii.** $m_i(\theta)$ forms a Donsker class. Hence

$$\mathbb{G}_n m_i(\theta) \Rightarrow \Delta(\theta) \text{ in } L^\infty(\Theta), \tag{F.6}$$

where $\Delta(\theta)$ is the Gaussian process with the covariance functions given in the statement of Theorem 2.4. Hence

$$\sup_{\theta \in \Theta} \|\mathbb{G}_n m_i(\theta)\| = O_p(1). \tag{F.7}$$

By condition **iv.**, uniformly in $\theta \in \Theta$ $W_n(\theta) = W(\theta) + o_p(1)$. Define $\xi_n \equiv \inf_{\theta \in \Theta}$ mineig $(W_n(\theta))$. By condition **iv.** $\xi_n \to_p \xi > 0$. Hence wp $\to 1$

$$\begin{aligned}
\inf_{v(\theta_I,\lambda) > K/\sqrt{n}} \left(\frac{\ell_n(\theta_I, \lambda)}{n(v(\theta_I, \lambda)^2 \wedge \delta^2)} \right) &\geq \inf_{v(\theta_I,\lambda) > K/\sqrt{n}} \xi_n \frac{\left(\|\sqrt{n}Em_i(\theta_I + \lambda/\sqrt{n}) + \mathbb{G}_n m_i(\theta_I + \lambda/\sqrt{n})\|^2 \right)}{n(v(\theta_I, \lambda)^2 \wedge \delta^2)} \\
&\geq \inf_{v(\theta_I,\lambda) > K/\sqrt{n}} \xi/2 \left\| \frac{\sqrt{n}Em_i(\theta_I + \lambda/\sqrt{n})}{\sqrt{n(v(\theta_I, \lambda)^2 \wedge \delta^2)}} + \frac{O_p(1)}{\sqrt{n(v(\theta_I, \lambda)^2 \wedge \delta^2)}} \right\|^2 \\
&\geq \inf_{v(\theta_I,\lambda) > K/\sqrt{n}} \xi/2 \left\| \frac{\sqrt{n}Em_i(\theta_I + \lambda/\sqrt{n})}{\sqrt{n(v(\theta_I, \lambda)^2 \wedge \delta^2)}} + \frac{O_p(1)}{\sqrt{n(v(\theta_I, \lambda)^2 \wedge \delta^2)}} \right\|_\infty^2
\end{aligned} \tag{F.8}$$

Note the line is different from the preceding one, as it uses the sup norm instead of the Euclidean norm.

By the partial identification condition (3.15)

$$\left\| \frac{\sqrt{n}Em_i(\theta_I + \lambda/\sqrt{n})}{\sqrt{n(v(\theta_I, \lambda)^2 \wedge \delta^2)}} \right\|_\infty > C, \tag{F.9}$$

for some constant $C > 0$. For a given $\epsilon > 0$, K can be made arbitrarily large, so that wp $\gtrsim 1 - \epsilon$

$$\left\| \frac{O_p(1)}{\sqrt{n(v(\theta_I, \lambda)^2 \wedge \delta^2)}} \right\|_\infty < C/2. \tag{F.10}$$

Hence wp $\gtrsim 1 - \epsilon$,

$$\inf_{v(\theta_I,\lambda) > K/\sqrt{n}} \left(\frac{\ell_n(\theta_I, \lambda)}{n(v(\theta_I, \lambda)^2 \wedge \delta^2)} \right) \geq C' = \xi/2(C/2)^2 \tag{F.11}$$

Hence wp $\gtrsim 1 - \epsilon$,

$$\ell_n(\theta_I, \lambda) \geq C' \cdot n \cdot (v(\theta_I, \lambda)^2 \wedge \delta^2) \quad \text{for all} \quad \nu(\theta_I, \lambda) > K/\sqrt{n}. \tag{F.12}$$

Hence C.3 is verified. □

Appendix G. Relationship to Pointwise Inference

Suppose that one is interested in a special parameter θ^* inside Θ_I. The inference about some θ^* in Θ_I is well motivated, when there is a sense in which θ^* is the "truth". This case typically arises in non-structural analysis or when it is believed that the models are correct representations of data-generating processes for some parameter value, i.e. there is $\theta^* \in \mathbb{R}^d$ such that the model law P_θ agrees with the actual stochastic law of data P. In this scenario, Θ_I is not of interest per se, but rather θ^* is. In method of moments settings, a similar problem has been already investigated in the context of dynamic model with censoring by Hu (2002), IV quantile estimation by Chernozhukov and Hansen (2003), and GMM weak or complete unidentification by Kleibergen (2002). Imbens and Manski (2004) investigate the Wald type inference about θ^* for the special case where a real parameter of interest is known to lie in an interval with endpoints that can be consistently estimated. Here we provide further insights concerning the pointwise inference in its relation to regionwise inference.

Assumption A.4. *Suppose there exists $a_n \to \infty$ such that*

$$a_n(Q_n(\theta_I) - q_n) \to_d \mathcal{C}(\theta_I) \quad \textit{for all} \quad \theta_I \in \Theta_I$$

where $\mathcal{C}(\theta_I)$ is a random variable. Moreover, for at least one $\theta_I \in \Theta_I$, $\mathcal{C}(\theta_I) > 0$ with positive probability and has continuous distribution function on $(0, \infty)$; otherwise, $\mathcal{C}(\theta_I) = 0$ with probability one.

Theorem G.1. *Suppose that Assumptions A.1 and A.4 hold. Let $c^*_\alpha = \sup_{\theta_I} c_\alpha(\theta_I)$. Then for any θ^* in Θ_I* $\liminf_{n\to\infty} P\Big\{\theta^* \in C_n(\hat{c}^*_\alpha)\Big\} \geq \alpha$.

Proof:

$$\begin{aligned}\liminf_{n\to\infty} P\{\theta^* \in C_n(c^*_\alpha)\} &= \liminf_{n\to\infty} P\{a_n\left(Q_n(\theta^*) - q_n\right) \leq \sup_{\theta_I} c_\alpha(\theta)\} \\ &\overset{(1)}{\geq} \liminf_{n\to\infty} P\{a_n\left(Q_n(\theta^*) - q_n\right) \leq c_\alpha(\theta^*)\} \geq (1 \text{ or } \alpha) \geq \alpha, \end{aligned} \tag{G.1}$$

□

The theorem above is constructed using the Anderson-Rubin pointwise testing principle. Notice that the confidence set constructed in the theorem above is necessarily smaller than the regionwise confidence set in Theorem 2.1. Notice also that as a byproduct of our derivation, inequality (1) in (G.3) shows that the set

$$C_n(c_\alpha(\cdot)) = \{\theta^* \in \widehat{\Theta}_I : a_n\left(Q_n(\theta^*) - q_n\right) \leq c_\alpha(\theta^*)\} \tag{G.2}$$

also has the α pointwise coverage. Hu (2001) proposed the set (G.2) in the context of a partially identified dynamic censored regression model. Chernozhukov and Hansen (2003) also use this technique for IV quantiles. Manski and Imbens (2004) construct this set for the case of interval-identified parameter.[13] The set $C_n(c_\alpha(\cdot))$ is generally not equal (is smaller) than the level set $C_n(c^*)$ of the function $a_nQ_n(\theta)$. However, this set is generally a *special case* of our construction. This can be seen by defining the new objective function $\widetilde{Q}_n(\theta) := Q_n(\theta)/\max[c_\alpha(\theta), \epsilon]$ for all $\theta \in \widehat{\Theta}_I$.[14]

Further let $\hat{c}_\alpha(\theta)$ be the subsampling estimate of $c_\alpha(\theta)$ for each $\theta \in \widehat{\Theta}_I$.

[13] A more recent paper than ours, by Ho, Pakes and Porter (2004), has also proposed such sets in the context of moment inequalities. The difference is that they do not directly work with objective functions.

[14] This is an equi-quantile transformation of the original objective function. In many examples this is unnecessary, as objective functions have the equi-quantile property by using optimal weights.

Theorem G.2. *Suppose that Assumptions A.1 and A.4 hold. Let* $\hat{c}^*_\alpha = \sup_{\theta_I} \hat{c}_\alpha(\theta_I)$. *Then for any* θ^* *in* Θ_I $\liminf_{n\to\infty} P\Big\{\theta^* \in C_n(\hat{c}^*_\alpha)\Big\} \geq \alpha$. *(Likewise, a corollary is that* $C_n(\hat{c}^*_\alpha(\cdot))$ *also covers* θ^* *with probability* α.*)*

Proof: Since $\Theta_I \subseteq \widehat{\Theta}_I$ wp $\to 1$, it follows that

$$
\begin{aligned}
\liminf_{n\to\infty} P\{\theta^* \in C_n(\hat{c}^*_\alpha)\} &= \liminf_{n\to\infty} P\{a_n\left(Q_n(\theta^*) - q_n\right) \leq \sup_{\theta_I} \hat{c}_\alpha(\theta)\} \\
&\overset{(1)}{\geq} \liminf_{n\to\infty} P\{a_n\left(Q_n(\theta^*) - q_n\right) \leq \hat{c}_\alpha(\theta^*)\} \\
&\overset{(2)}{\geq} \liminf_{n\to\infty} P\{a_n\left(Q_n(\theta^*) - q_n\right) \leq c_\alpha(\theta^*) + o_p(1)\} \\
&\geq (1 \text{ or } \alpha) \geq \alpha,
\end{aligned} \tag{G.3}
$$

Equality (2) follows from the standard argument for subsampling, e.g. as the one presented in Step 2 of the proof of Theorem 2.1. (Inequality (1) shows that $C_n(\hat{c}^*_\alpha(\cdot))$ also covers θ^* with probability α.).

□

[CONSISTENCY AND RATES HERE TOO?]

Notation and Terms

$\to_p$	convergence in (outer) probability P^*
$\to_d$	convergence in distribution under P^*
wp $\to 1$	with inner probability P_* converging to one
wp $\gtrsim 1-\epsilon$	with inner probability P_* larger than $1-\epsilon$ for sufficiently large n,
$B_\delta(x)$	closed ball centered at x of radius $\delta > 0$
I	identity matrix
$\mathcal{N}(0, a)$	normal random vector with mean 0 and variance matrix a
$\mathcal{F}$ Donsker class	here this means that empirical process $f \mapsto \frac{1}{\sqrt{n}}\sum_{i=1}^n (f(W_i) - Ef(W_i))$ is asymptotically Gaussian in $L^\infty(\mathcal{F})$, see Vaart (1999)
$L^\infty(\mathcal{F})$	metric space of bounded over $\mathcal{F}$ functions, see Vaart (1999)
mineig(A)	minimum eigenvalue of matrix A

References

Andrews, D. W. K. (1999): "Estimation when a parameter is on a boundary," *Econometrica*, 67(6), 1341–1383.

——— (2000): "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space," *Econometrica*, 68(2), 399–405.

Angrist, J., and A. Krueger (1992): "Age at School Entry," *JASA*, 57, 11–25.

Chernozhukov, V., and C. Hansen (2001): "An IV Model of Quantile Treatment Effects," forthcoming in *Econometrica.*

Chernozhukov, V., and H. Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.

Ciliberto, F., and E. Tamer (2003): "Market Structure and Multiple Equilibria in Airline Markets," Working Paper: Princeton University.

Demidenko, E. (2000): "Is this the least squares estimate?," *Biometrika*, 87(2), 437–452.

Gisltein, C. Z., and E. Leamer (1983): "Robust Sets of Regression Estimates," *Econometrica*, 51.

Hansen, L., J. Heaton, and E. G. Luttmer (1995): "Econometric Evaluation of Asset Pricing Models," *The Review of Finacial Studies*, 8, 237–274.

HOROWITZ, J., AND C. MANSKI (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281–302.

——— (1998): "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37–58.

HU, L. (2002): "Estimation of a Censored Dynamic Panel Data Model," *Econometrica*, 70(6).

IMBENS, G., AND C. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," Department of Economics, Northwestern University.

KLEIBERGEN, F. (2002): "Pivotal statistics for testing structural parameters in instrumental variables regression," *Econometrica*, 70(5), 1781–1803.

KLEPPER, S., AND E. LEAMER (1984): "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica*, 52(1), 163–184.

MANSKI, C. (2003): *Partial Identification of Probability Distributions*, Springer Series in Statistics. Springer, New York.

MANSKI, C., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–547.

MARSCHAK, J., AND W. H. ANDREWS (1944): "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 812, 143–203.

MILNOR, J. (1964): "Differential topology," in *Lectures on Modern Mathematics, Vol. II*, pp. 165–183. Wiley, New York.

NERLOVE, M. (1965): *Estimation and Identification of Cobb-Douglas Production Functions*. Rand McNally & and Company, Chicago.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," in *Handbook of econometrics, Vol. IV*, vol. 2 of *Handbooks in Econom.*, pp. 2111–2245. North-Holland, Amsterdam.

ROBERT, C. P., AND G. CASELLA (1998): *Monte Carlo Statistical Methods*. Springer-Verlag.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak convergence and empirical processes*. Springer-Verlag, New York.

POLITIS, D., J. ROMANO, AND M. WOLF (1999): *Subsampling*, Springer Series in Statistics. Springer:New York.

VAART, A. V. D. (1999): *Asymptotics Statistics*. Cambridge University Press.