A Dual Approach to Inference for Partially Identified Econometric Models

Hiroaki Kaido*
Department of Economics
Boston University

December, 2015

Abstract

This paper considers inference for the set Θ_I of parameter values that minimize a criterion function. Chernozhukov, Hong, and Tamer (2007) (CHT) develop a general theory of consistent set estimation using the level-set of a criterion function and inference based on a quasi-likelihood ratio (QLR)-type statistic. This paper establishes a dual relationship between the level-set estimator and its support function and shows that the properly normalized (scaled and centered) support function provides an alternative Wald-type inference method to conduct tests regarding the identified set and a point θ_0 in the identified set. These tests can be inverted to obtain confidence sets for Θ_I and θ_0. For econometric models that involve finitely many moment inequalities, we show that our Wald-type statistic is asymptotically equivalent to CHT's QLR statistic under regularity conditions.

JEL Classification: C12

Keywords: *Partial Identification, Criterion Function, Support Function*

*I thank anonymous referees for helpful comments that have helped improve the paper. I am deeply grateful to Halbert White for his guidance and constant encouragement. I am also indebted to Jin Seo Cho, Graham Elliott, Ben Gillen, Patrik Guggenberger, Philip Haile, Joel Horowitz, Ivana Komunjer, Charles Manski, Francesca Molinari, Jack Porter, Andres Santos, Max Stinchcombe, Yixiao Sun, Jörg Stoye, Allan Timmermann, and numerous seminar participants. I gratefully acknowledge financial support from the Nakajima Foundation. Email: hkaido@bu.edu

1 Introduction

Statistical inference for partially identified economic models is a growing field in econometrics. The field was pioneered by Charles Manski in the 1990's (See Manski, 2003, and the references therein), and there have since been substantial theoretical extensions and applications. In this literature, the economic structures of interest are characterized by an *identified set* Θ_I, rather than by a single point in the parameter space $\Theta \subset \mathbb{R}^d, d \in \mathbb{N}$. Elements of the identified set lead to observationally equivalent data generating processes. A sample of data generated by any of the parameter values in the identified set, therefore, gives us information about the identified set, but not about the underlying "true" parameter value generating the observed data.

Chernozhukov, Hong, and Tamer (2007) (CHT) study estimation and statistical inference on Θ_I within a general extremum estimation framework. CHT have shown that a level-set estimator based on a properly chosen sequence of levels for the criterion function consistently estimates the identified set, defined as a set of minimizers. They use a quasi-likelihood ratio (QLR) statistic to construct a confidence set that asymptotically covers the identified set with at least a prespecified probability. This criterion function approach is applicable to a broad class of problems.

Another common approach is to estimate the boundary of Θ_I directly. This is an attractive alternative if the boundary of the identified set is easily estimable. Recent studies show that when Θ_I is a compact convex set, its support function provides a tractable representation by summarizing the location of the supporting hyperplanes of Θ_I. (Beresteanu and Molinari, 2008 (BM); Bontemps, Magnac, and Maurin, 2012). So far, the criterion function approach and the support function approach have been viewed as distinct. Each has its advantages and challenges. The criterion function approach is widely applicable, but constructing the level set can be computationally demanding. The support function approach, on the other hand, is more direct and computationally tractable for some problems, but it has been applied to a limited class of models when parameters are multi-dimensional. A main contribution of this paper is to unify these approaches within a general framework. We do this by studying an inference method that is based on the support function of a level set estimator. To the best of our knowledge, this is the first such effort.

In this paper, we focus on econometric models with compact convex identified sets, which enables us to characterize the identified set by its support function[1]. This class includes many econometric models studied recently, e.g., regression with interval data (Manski and Tamer, 2002; Magnac and Maurin, 2008), a class of discrete choice models Pakes (2010), consumer demand models with unobserved heterogeneity (Blundell, Kristensen, and Matzkin, 2014), and an asset pricing model in incomplete markets (Kaido and White, 2009). Following CHT, our estimator of Θ_I is the level set $\hat{\Theta}_n = \{\theta : Q_n(\theta) \leq t_n\}$ of a criterion function $Q_n(\cdot)$ for some sequence of levels $\{t_n\}$. The support function approach provides a straightforward algorithm to compute the boundary of this estimator. Specifically, we propose to solve the optimization problem $\max_{Q_n(\theta)\leq t_n}\langle p, \theta\rangle$ for each p. This yields

[1]Our analysis applies to the convex hull of the identified set if it is nonconvex.

the support function $s(\cdot, \hat{\Theta}_n)$ of the set estimator as a value function and also gives the boundary of $\hat{\Theta}_n$. The optimization is a convex programming problem, which can be solved using standard algorithms.

The estimated support function can also be used to conduct inference. Using a dual relationship between the criterion function and support function, we first show that the asymptotic distribution of the properly normalized (centered and scaled) support function is that of a specific stochastic process on the unit sphere. The normalized support function lets us make various types of inference for Θ_I and points in Θ_I. For example, as shown in BM, the normalized support function allows one to construct a confidence set that covers the identified set with at least some prescribed confidence level. Further, one may test whether Θ_I includes a specific point, i.e., $H_0 : \theta_0 \in \Theta_I$ using a test statistic based on the estimated support function. We contribute to the literature by establishing the asymptotic distribution of this statistic. Specifically, our asymptotic distribution result generally holds even if the identified set has kink points and thus extends the result of Bontemps, Magnac, and Maurin (2012). This test can be inverted to construct a confidence set for each point in the identified set.

Our work is related to the work of BM who first studied inference based on estimated support functions for the case where Θ_I is a linear transformation of the Aumann expectation of set-valued random variables and Bontemps, Magnac, and Maurin (2012) who consider a confidence set for a point in the identified set, when Θ_I is characterized by incomplete linear moment restrictions. Our analysis further contributes to this line of research by extending these results to the general setting where Θ_I is the set of minimizers of a convex criterion function.

We apply the main results to econometric models characterized by finitely many moment inequalities. This class has been extensively studied recently (see references in Section 4). We contribute to this literature by establishing a new equivalence result within this class. Our Wald-type statistic (squared directed Hausdorff distance) and CHT's QLR statistic converge in distribution to the same limit under some regularity conditions. As a result, the Wald confidence set, a set obtained by expanding the set estimator by a suitable critical value, is asymptotically equivalent to CHT's confidence set, a level set whose level is a specific quantile of the QLR statistic.

The paper is organized as follows. In section 2, we summarize CHT's econometric framework and introduce some useful background. We establish the asymptotic distribution of the normalized support function and develop our inference methods in section 3. Section 4 studies moment inequality models. We present Monte Carlo simulation results in section 5 and conclude in section 6. We collect our mathematical proofs in the appendix.

Throughout, we use the following notation. Let $\mathbb{R}_+ := [0, \infty)$ and $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$. For any closed set $A \subseteq \mathbb{R}^d$, let ∂A denote its boundary, and let A^o denote its interior. For any $x, y \in \mathbb{R}^d$, let $\langle x, y \rangle$ denote the inner product of x and y, and let $\|x\|$ denote the Euclidean norm of x. We let $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ denote the unit sphere in $\mathbb{R}^d$, and $\mathcal{C}(\mathbb{S}^{d-1})$ is the set of continuous functions on $\mathbb{S}^{d-1}$. Finally, for any $J \times J$ matrix w and vector $y \in \mathbb{R}^J$, we let

$\|wy\|_+ := \|w(y \circ 1\{y \geq 0\})\|$, where $\circ$ denotes the entrywise product.

2 General Setup

2.1 Criterion Functions and Set Estimator

We start with introducing criterion functions and high level conditions (Assumptions 2.1-2.3) based on the conditions in CHT. Our first assumption is on the data generating process (DGP), parameter space, and the criterion functions.

Assumption 2.1. *(i) Let $(\Omega, \mathfrak{F}, P)$ be a complete probability space. Let $d \in \mathbb{N}$, and let $\Theta \subseteq \mathbb{R}^d$ be a compact and convex parameter space with a nonempty interior; (ii) Let $Q : \mathbb{R}^d \to \bar{\mathbb{R}}_+$ be a lower semicontinuous (lsc) function; (iii) For $n = 1, 2, \ldots$, let $Q_n : \Omega \times \mathbb{R}^d \to \bar{\mathbb{R}}_+$ be a jointly measurable function such that $Q_n(\omega, \theta) < \infty$ for at least one $\theta \in \Theta$, $Q_n(\omega, \theta) = \infty$ for all $\theta \notin \Theta$, and $\theta \mapsto Q_n(\omega, \theta)$ is lsc with probability 1.*

Compactness is a standard assumption on Θ for extremum estimation. The function Q_n acts as our sample criterion function. For example, a commonly used criterion function for moment inequality models is

$$Q_n(\omega, \theta) = \|\hat{W}_n^{1/2}(\omega, \theta)\frac{1}{n}\sum_{i=1}^{n} m(X_i(\omega), \theta)\|_+^2, \tag{2.1}$$

where $m(x, \theta)$ is a vector-valued function such that $E[m(X_i, \theta)] \leq 0$ for one or more values of θ, and $\hat{W}_n$ is a weighting matrix that can depend on the sample. For simplicity, we write $Q_n(\theta)$ below, but its dependence on ω should be understood implicitly. The function Q is the population criterion function. Without loss of generality, we normalize the minimum value of Q to 0. Following CHT, we then define the identified set as the set of minimizers of Q:

$$\Theta_I := \{\theta \in \Theta : Q(\theta) = 0\}. \tag{2.2}$$

Throughout, we assume that Θ_I is a non-empty subset of Θ. The set estimator of Θ_I is then defined as a level-set of Q_n. We also normalize Q_n so that the minimum of Q_n is 0. For a non-negative sequence $\{t_n\} \subset \mathbb{R}_+$ and a positive sequence $\{a_n\} \subset \mathbb{R}_+$, the *set estimator* is defined by

$$\hat{\Theta}_n(t_n) := \{\theta \in \Theta : a_n Q_n(\theta) \leq t_n\}. \tag{2.3}$$

For any $a \in \mathbb{R}^d$ and closed set $B \subseteq \mathbb{R}^d$, let $d(a, B) := \inf_{b \in B} \|a - b\|$. For any closed subsets A, B of $\mathbb{R}^d$, let

$$d_H(A, B) := \max\left[\vec{d}_H(A, B), \vec{d}_H(B, A)\right], \quad \vec{d}_H(A, B) := \sup_{a \in A} d(a, B)\ , \tag{2.4}$$

where d_H and $\vec{d}_H$ are the Hausdorff and directed Hausdorff distances respectively. The following assumptions based on CHT's conditions C.1-C.3 are general enough to be satisfied by many examples involving inequality constraints.

Assumption 2.2. *(i)* $\sup_{\theta\in\Theta}\{Q(\theta) - Q_n(\theta)\}_+ = o_p(1)$. *(ii)* $\sup_{\theta\in\Theta_I} Q_n(\theta) = O_p(1/a_n)$. *(iii) There exist positive constants* (δ, κ, γ) *such that for any* $\epsilon \in (0,1)$, *there are* $(\kappa_\epsilon, n_\epsilon)$ *such that for all* $n \geq n_\epsilon$

$$Q_n(\theta) \geq \kappa \min\{d(\theta, \Theta_I), \delta\}^\gamma,$$

uniformly on $\{\theta \in \Theta : d(\theta, \Theta_I) \geq (\kappa_\epsilon/a_n)^{1/\gamma}\}$ *with probability at least* $1 - \epsilon$.

Assumption 2.3 (Degeneracy). *(i) There is a sequence of subsets* Θ_n *of* Θ, *which could be data dependent such that* Q_n *vanishes on these subsets, that is,* $Q_n(\theta) = 0$ *for each* $\theta \in \Theta_n$, *for each* n, *and these sets can approximate the identified set arbitrarily well in the Hausdorff metric, that is,* $d_H(\Theta_n, \Theta_I) \leq \epsilon_n$ *for some* $\epsilon_n = o_p(1)$. *(ii)* $\epsilon_n = O_p(a_n^{-1/\gamma})$.

Under Assumptions 2.1-2.3, CHT's Theorem 3.2 is applicable. In particular, CHT show that it is possible to achieve consistency and an exact polynomial rate of convergence by choosing a level $t_n = t \in \mathbb{R}_+$ such that $t \geq \inf_{\theta\in\Theta} a_n Q_n(\theta)$ with probability 1. Hence, we have $d_H(\hat{\Theta}_n(t), \Theta_I) = O_p(a_n^{-1/\gamma})$.

Finally, we assume that Θ_I and $\hat{\Theta}_n(t)$ are convex by assuming that the population and sample criterion functions are convex.

Assumption 2.4. *(i)* Q *is a convex function; (ii)* Q_n *is a convex function a.s.*

2.2 Support Function

Throughout, we use support functions to characterize compact convex sets. The *support function* $s(\cdot, F) : \mathbb{S}^{d-1} \to \mathbb{R}$ of a compact convex set $F \subset \mathbb{R}^d$ is defined pointwise by

$$s(p, F) = \sup_{x\in F}\langle p, x\rangle. \tag{2.5}$$

The set $H(p, F) = \{x \in \mathbb{R}^d : \langle p, x\rangle = s(p, F)\} \cap F$ is called the *support set*. Heuristically, for each unit vector $p \in \mathbb{S}^{d-1}$, the support function $s(p, F)$ measures the signed distance from the origin of the supporting plane of the set F with a normal vector p (see the online addendum for more detailed descriptions of these objects). The support function is a continuous function on the unit sphere and therefore takes its value in $\mathcal{C}(\mathbb{S}^{d-1})$. Since any convex set can be represented by the intersection of such supporting planes, the support function fully characterizes the boundary of the set of interest.

In our setting, the support function of the set estimator $\hat{\Theta}_n(t)$ offers a straightforward procedure to compute the boundary of $\hat{\Theta}_n(t)$. Consider the following optimization problem:

$$\text{maximize} \quad \langle p, \theta\rangle, \quad \text{subject to} \quad a_n Q_n(\theta) \leq t. \tag{2.6}$$

The optimal value function of this problem is $s(p, \hat{\Theta}_n(t))$, and a solution to (2.6) is a point in the support set $H(p, \hat{\Theta}_n(t))$. One may then trace out the boundary of $\hat{\Theta}_n(t)$ by solving (2.6) for different values of p. The optimization problem in (2.6) is a convex programming problem, which is often easily solvable using standard algorithms (see for example Boyd and Vandenberghe, 2004).

In addition to providing a straightforward algorithm to compute $\hat{\Theta}_n(t)$, support functions have useful properties for inference. That is, for any compact convex sets A, B, it holds that $d_H(A,B) = \sup_{p\in\mathbb{S}^{d-1}} |s(p,A) - s(p,B)|$ and $\vec{d}_H(A,B) = \sup_{p\in\mathbb{S}^{d-1}} \{s(p,A) - s(p,B)\}_+$.[2] This means that the Hausdorff distance (or the directed Hausdorff distance) between sets is equal to the uniform distance (or the one-sided uniform distance) between the support functions. These isometry relationships between sets and support functions allow us to write

$$a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) = \sup_{p\in\mathbb{S}^{d-1}} |\mathcal{Z}_n(p,t)|, \quad \text{and} \quad a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) = \sup_{p\in\mathbb{S}^{d-1}} \{-\mathcal{Z}_n(p,t)\}_+, \tag{2.7}$$

where $\mathcal{Z}_n(p,t)$ is the normalized support function defined by

$$\mathcal{Z}_n(p,t) := a_n^{1/\gamma}\big(s(p,\hat{\Theta}_n(t)) - s(p,\Theta_I)\big). \tag{2.8}$$

Therefore, if for a given t, $\mathcal{Z}_n(\cdot,t)$ converges weakly to some limit $\mathcal{Z}(\cdot,t)$ in $\mathcal{C}(\mathbb{S}^{d-1})$, then the desired limiting distributions of Hausdorff distance measures follow from the continuous mapping theorem. This in turn allows us to conduct inference for Θ_I and points inside it.

2.3 Examples

To fix ideas, we discuss below leading examples of models with convex identified sets based on simplifications of well known models. The first example is a regression model with an interval-valued outcome studied in Manski and Tamer (2002).

Example 1 (Interval censored outcome)**.** An outcome variable is generated as

$$Y = Z'\theta + \epsilon,$$

where $Z \in \mathbb{R}^d$ is a regressor vector with discrete support $\mathcal{Z} \equiv \{z_1, \cdots, z_J\}$, $\theta \in \Theta \subseteq \mathbb{R}^d$, and $E[\epsilon|Z] = 0$. Y is not observed but the outcome interval $[Y_L, Y_U]$ which contains Y is observed. The identified set then consists of parameter values that satisfy

$$E[Y_L|Z = z_j] \le z_j'\theta \le E[Y_U|Z = z_j], \quad j = 1, \cdots, K. \tag{2.9}$$

Since the constraints are affine in θ, the identified set is convex. Let $1_{\mathcal{Z}}(z) = (1\{z = z_1\}, \cdots, 1\{z = z_K\})'$ and $m(x) \equiv (y_L 1_{\mathcal{Z}}(z), y_U 1_{\mathcal{Z}}(z), 1_{\mathcal{Z}}(z))$. Further, let $A \equiv (-z_1, \cdots, -z_K, z_1, \cdots, z_K)'$. The affine constraints in (2.9) can then be written as

$$A\theta - F(E[m(X)]) \le 0\ , \tag{2.10}$$

where $F : \mathbb{R}^{3K} \to \mathbb{R}^{2K}$ is a transformation that combines unconditional moments to construct

[2]See Theorem 1.1.12 (Hörmander's embedding theorem) in Li, Ogura, and Kreinovich (2002) and Lemma A.1 in BM.

conditional moments whose k-th component is defined as follows:

$$F_k(v) = \begin{cases} \frac{v_k}{v_{2K+k}}, & k = 1, \cdots, K \\ -\frac{v_k}{v_{2K+k}}, & k = K+1, \cdots, 2K. \end{cases} \tag{2.11}$$

Other examples that give the constraints of the form in (2.10) include the IV model for a binary outcome studied in Chesher (2009) and a special case of revealed preference bounds studied in Blundell, Kristensen, and Matzkin (2014).

Strictly convex identified sets also arise in an asset pricing model with market frictions.

Example 2 (Pricing kernel)**.** Let $Z : \Omega \to \mathbb{R}^J$ be the payoffs of J securities that are traded at a price of $V \in \mathbb{R}^J_+$. If short sales are not allowed for any securities, then the feasible set of portfolio weights is restricted to $\mathbb{R}^J_+$ and the standard Euler equation does not hold. Instead, under power utility, the following Euler inequalities hold (see Luttmer, 1996):

$$E[\frac{1}{1+\rho} Y^{-\gamma} Z - V] \leq 0, \tag{2.12}$$

where $Y : \Omega \to \mathbb{R}_+$ is a state variable, e.g. consumption growth, ρ is the investor's subjective discount rate, and γ is the relative risk aversion coefficient. When the payoff Z takes nonnegative values almost surely, the set of parameter values $\theta = (\rho, \gamma)$ that satisfy (2.12) is convex. A criterion function can then be defined as in (2.1) with $X = (Y, Z', V)'$ and $m_j(x, \theta) = \frac{1}{1+\rho} y^{-\gamma} z - v$ for all j. This example belongs to the class of moment inequality models studied in Section 4.1.

3 Inference

We first establish the main duality that relates the stochastic behavior of the normalized support function $\mathcal{Z}_n(\cdot, t)$ to that of a localized criterion function. This result is then used to show that $\mathcal{Z}_n(\cdot, t)$ converges weakly to a limit $\mathcal{Z}(\cdot, t)$, which in turn ensures the asymptotic validity of Wald-type inference methods based on the normalized support function.

3.1 Duality and the Asymptotic Distribution of $\mathcal{Z}_n$

To see how the criterion function and the support function of the set estimator are related to each other, we start with the following equivalence relationship:

$$s(p, \hat{\Theta}_n(t)) < u \quad \Leftrightarrow \quad \inf_{\theta \in K_{u,p} \cap \Theta} a_n Q_n(\theta) > t, \ a.s., \tag{3.1}$$

where $K_{u,p} := \{\theta \in \mathbb{R}^d : \langle p, \theta \rangle \geq u\}$.[3] The left hand side of (3.1) means the set estimator is separated from the half-space $K_{u,p}$ intersected with Θ. Since $\hat{\Theta}_n(t)$ is the t-level set of $a_n Q_n$, this

[3]The proof of duality results discussed here are collected in Appendix C. Note that if $\hat{\Theta}_n(t) = \varnothing$, we take $s(p, \hat{\Theta}_n(t)) = \sup_{\theta \in \varnothing} \langle p, \theta \rangle = -\infty$.

means that the minimum value of the function over $K_{u,p} \cap \Theta$ exceeds the chosen level t. Figure 1 illustrates this relationship.

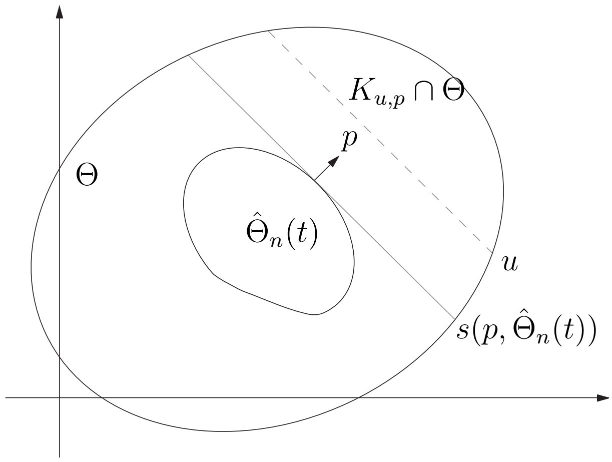

Figure 1: Set estimator $\hat{\Theta}_n(t)$ and its support function $s(p, \hat{\Theta}_n(t))$

A key object for our asymptotic analysis is the sample criterion function localized on a neighborhood of Θ_I. For each $(\theta, \lambda) \in \partial\Theta_I \times \mathbb{R}^d$, define the *local criterion function* ℓ_n by

$$\ell_n(\theta, \lambda) := a_n Q_n(\theta + \lambda / a_n^{1/\gamma}).$$

A dual relationship similar to (3.1) also holds for the local criterion function ℓ_n and the normalized support function $\mathcal{Z}_n$. In particular, using (3.1), Lemma C.2 in the appendix shows that the following relationship holds:

$$\mathcal{Z}_n(p, t) < u \quad \Leftrightarrow \quad \inf_{R_{n,u,p}} \ell_n(\theta, \lambda) > t, \;\; \forall (p, u) \in \mathbb{S}^{d-1} \times \mathbb{R}, \;\; a.s., \tag{3.2}$$

where $R_{n,u,p} := \{(\theta, \lambda) : \theta \in H(p, \Theta_I), \lambda \in K_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta)\}$ consists of the values of (θ, λ) such that θ is in the support set $H(p, \Theta_I)$ and λ is in the local parameter space $K_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta)$.

The equivalence relationship in (3.2) implies that the distribution of $\mathcal{Z}_n(\cdot, t)$ is tied to that of the infimum of ℓ_n over the set $R_{n,u,p}$. The finite dimensional convergence of $\mathcal{Z}_n(\cdot, t)$ is then ensured if the probability of the event on the right hand side of (3.2) converges properly to some limit. Heuristically, the argument for establishing the limiting distribution can be summarized as follows. For any $\{(p_k, u_k)\}_{k=1}^m$, the duality in (3.2) implies

$$P(\mathcal{Z}_n(p_1, t) < u_1, \cdots, \mathcal{Z}_n(p_m, t) < u_m) = P\Big(\inf_{R_{n,u_1,p_1}} \ell_n(\theta, \lambda) > t, \cdots, \inf_{R_{n,u_m,p_m}} \ell_n(\theta, \lambda) > t\Big). \tag{3.3}$$

If $\{\inf_{R_{n,u_k,p_k}} \ell_n(\theta, \lambda)\}_{k=1}^m$ converges in distribution to $\{\inf_{R_{u_k,p_k}} \ell_\infty(\theta, \lambda)\}_{k=1}^m$ for some process ℓ_∞ and suitable sets $\{R_{u_k,p_k}\}_{k=1}^m$, we can seek a process $\mathcal{Z}$ such that

$$P(\mathcal{Z}(p_1, t) < u_1, \cdots, \mathcal{Z}(p_m, t) < u_m) = P\Big(\inf_{R_{u_1,p_1}} \ell_\infty(\theta, \lambda) > t, \cdots, \inf_{R_{u_m,p_m}} \ell_\infty(\theta, \lambda) > t\Big). \tag{3.4}$$

Then, $\mathcal{Z}_n(\cdot, t)$ converges weakly in finite dimension to $\mathcal{Z}(\cdot, t)$. The next theorem establishes this; It further gives the asymptotic distributions of the Hausdorff distances by showing that $\mathcal{Z}_n(\cdot, t)$

converges weakly to $\mathcal{Z}(\cdot,t)$ in $\mathcal{C}(\mathbb{S}^{d-1})$. Below, we use $\stackrel{u.d.}{\to}$ to denote weak convergence in $\mathcal{C}(\mathbb{S}^{d-1})$.[4]

Theorem 3.1. *Suppose that Assumptions 2.1-2.4, and B.1 (in the appendix) hold. Then, (i) for each $t \in \mathbb{R}_+$, $\mathcal{Z}_n(\cdot,t)$ converges weakly in finite dimension to $\mathcal{Z}(\cdot,t)$, where $\mathcal{Z}(\cdot,t)$ is a stochastic process on $\mathbb{S}^{d-1}$, which has the representation:*

$$\mathcal{Z}(p,t) = \sup_{\theta \in H(p,\Theta_I)} s\big(p, \Lambda_{\theta,t}\big) \ , \quad \Lambda_{\theta,t} = \{\lambda : \ell_\infty(\theta,\lambda) \le t\} \ , \tag{3.5}$$

where ℓ_∞ is defined in Assumption B.1 in the appendix; (ii) Further, $\mathcal{Z}_n(\cdot,t) \stackrel{u.d.}{\to} \mathcal{Z}(\cdot,t)$ so that

$$a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \stackrel{d}{\to} \sup_{p\in\mathbb{S}^{d-1}} |\mathcal{Z}(p,t)| \ , \quad \text{and} \quad a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \stackrel{d}{\to} \sup_{p\in\mathbb{S}^{d-1}} \{-\mathcal{Z}(p,t)\}_+ . \tag{3.6}$$

Theorem 3.1 characterizes the limiting distribution of the support function of the set estimator and provides a basis for asymptotically valid inference. We illustrate this result using Example 1.

Example 1 (Interval censored outcome (continued))**.** Recall that the inequality restrictions in (2.9) can be written as constraints affine in θ. Using this, define a sample criterion function Q_n by

$$Q_n(\theta) = \sum_{k=1}^{2K} \hat{\sigma}_{k,n}^{-1} \Big(a_k'\theta - F_k(\hat{E}_n[m(X_i)]\Big)_+ , \tag{3.7}$$

where a_k denotes the k-th row of A in (2.10), and $\hat{\sigma}^2_{k,n}$ is a consistent estimator of the asymptotic variance of the k-th constraint.[5] One can then use the support function of a level-t set $\hat{\Theta}_n(t)$ of this criterion function for inference. We will discuss how to compute $\hat{\Theta}_n(t)$ and a confidence region using convex programs in the next section.

By the dual relationship in (3.2), the asymptotic behavior of the normalized support function $\mathcal{Z}_n(p,t)$ is tied to that of the following local criterion function:

$$\begin{aligned} \ell_n(\theta,\lambda) &= \sqrt{n} \sum_{k=1}^{2K} \hat{\sigma}_{k,n}^{-1} \Big(a_k'(\theta + \lambda/\sqrt{n}) - F_k(\hat{E}_n[m(X_i)])\Big)_+ \\ &= \sum_{k=1}^{2K} \hat{\sigma}_{k,n}^{-1} \Big(\sqrt{n}\{F_k(E[m(X_i)]) - F_k(\hat{E}_n[m(X_i)])\} + a_k'\lambda + \sqrt{n}\{a_k'\theta - F_k(E[m(X_i)])\}\Big)_+ , \end{aligned} \tag{3.8}$$

where $m(x) = (y_L 1_{\mathcal{Z}}(z), y_U 1_{\mathcal{Z}}(z), 1_{\mathcal{Z}}(z))$. This local criterion function $\ell_n(\theta,\lambda)$ can then be shown to converge in the mode required by Theorem 3.1 to the following limiting process:

$$\ell_\infty(\theta,\lambda) = \sum_{k=1}^{2K} \sigma_k^{-1} \Big(\mathbb{G}_k + a_k'\lambda + \varsigma_k(\theta)\Big)_+ \ , \tag{3.9}$$

where $\mathbb{G} \in \mathbb{R}^{2K}$ is a multivariate normal vector with the covariance matrix $\nabla F \Omega_m \nabla F'$ where ∇F is the gradient of F, Ω_m is the covariance matrix of $m(X_i)$, and $\varsigma_k(\theta) = 0$ if $a_k'\theta - F_k(E[m(X_i)]) = 0$,

[4] The online addendum (`http://people.bu.edu/hkaido/pdf/Supp_Duality.pdf`) provides discussions on the convergence modes of stochastic processes used in this paper.

[5] Due to the moments being affine in θ and A being known, the asymptotic variance of the constraints does not depend on θ in this example.

and $\varsigma_j(\theta) = -\infty$ otherwise.[6] By Theorem 3.1, the limiting distribution of the normalized support function $\mathcal{Z}_n(p,t)$ depends on ℓ_∞ via (3.5), which governs the asymptotic properties of Wald-type statistics we introduce below. Theorem 3.1 therefore gives a theoretical basis for the asymptotic validity of our inference methods.

3.2 Inference

For making asymptotically valid inference, one needs to consistently estimate critical values of the form:

$$c_{1-\alpha}(t) := \inf\Big\{x : P\Big(\sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}(p,t)) \le x\Big) \ge 1-\alpha\Big\}, \tag{3.10}$$

where $\Psi_0 \subseteq \mathbb{S}^{d-1}$, and $\Upsilon : \mathbb{R} \to \mathbb{R}$ is a known function. Below, we assume Q_n is constructed from a sample $\{X_i : \Omega \to \mathbb{R}^k\}_{i=1}^n$ of IID random vectors and give a generic subsampling procedure for estimating $c_{1-\alpha}$.

Assumption 3.1. *Let Assumption 2.1 hold with $Q_n(\omega,\theta) = \tilde{Q}_n(X_1(\omega),\cdots,X_n(\omega),\theta)$ where $\tilde{Q}_n : \prod_{i=1}^n \mathbb{R}^k \times \mathbb{R}^d \to \bar{\mathbb{R}}_+$ is jointly measurable, $n = 1,2,\cdots$, and $\{X_i\}$ is an IID sequence of random k-vectors, $k \in \mathbb{N}$.*

Under Assumption 3.1, a straightforward subsampling algorithm is the following.

Algorithm 3.1 (Subsampling for normalized support functions)**.** Let $t > 0$ and $0 < \alpha < 1$ be given. Let $b := b_n < n$ be a positive integer, and let $N_{n,b} := \binom{n}{b}$. Let $\{\Psi_n\}$ be a sequence of random closed subsets of $\mathbb{S}^{d-1}$.

Step 1. For $k = 1,\cdots,N_{n,b}$, construct $\hat{\Theta}_{n,b,k}(t)$, the set estimator for the k-th subsample, computed as a t-level set of the criterion function $a_b\tilde{Q}_{n,b,k}(X_{k_1},\cdots,X_{k_b},\theta)$.

Step 2. For $k = 1,\cdots,N_{n,b}$, compute $\mathcal{Z}_{n,b,k}(p,t) := a_b^{1/\gamma}[s(p,\hat{\Theta}_{n,b,k}(t)) - s(p,\hat{\Theta}_n(t))]$.

Step 3. Compute the $(1-\alpha)$-quantile $\hat{c}_{n,b,1-\alpha}(t)$ of the subsampling distribution:

$$\hat{F}_{n,b}(x,t) := N_{n,b}^{-1} \sum_{1\le k\le N_{n,b}} 1\Big\{\sup_{p\in\Psi_n} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) \le x\Big\}. \tag{3.11}$$

For any t, let $F(x,t) := P(\sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}(p,t)) \le x)$. The next theorem is a basic result for subsampling statistics based on the normalized support function.

Theorem 3.2. *Suppose the conditions of Theorem 3.1 and Assumption 3.1 hold. Suppose further that Υ is Lipschitz continuous, Ψ_0 is compact, and that $d_H(\Psi_n,\Psi_0) = o_p(1)$. Let $\hat{F}_{n,b}(\cdot,t)$ and $\hat{c}_{n,b,1-\alpha}(t)$ be computed by Algorithm 3.1. Suppose that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. If x is a continuity point of $F(\cdot,t)$, then $\hat{F}_{n,b}(x,t) \to F(x,t)$ in probability;*

[6] The analysis of this example is similar to that for the moment inequalities, which we will discuss in detail in Section 4.1. A difference is due to the presence of the transformation F.

Remark 3.1. Subsampling is generally valid under Assumption 3.1 and the conditions of Theorem 3.1. We note, however, that if the example of interest has additional structure, an alternative inference method may be preferable in terms of the accuracy of approximation or computational tractability. For example, the score-based weighted bootstrap applied to models defined by convex moment inequalities in Kaido and Santos (2014) does not require repeated set estimation on bootstrap samples and is therefore computationally more efficient.

3.2.1 Inference for the Identified Set

We illustrate the use of Theorem 3.1 by studying inference for the identified set. Let Θ_0 be a compact convex set, and consider testing

$$H_0 : \Theta_0 \subseteq \Theta_I \quad vs. \quad H_1 : \Theta_0 \nsubseteq \Theta_I. \tag{3.12}$$

We test this hypothesis using the scaled directed Hausdorff distance:

$$T_n^{\rightarrow}(t) := a_n^{1/\gamma} \vec{d}_H(\Theta_0, \hat{\Theta}_n(t)). \tag{3.13}$$

Theorem 3.1 shows that under the null hypothesis, $T_n^{\rightarrow}$ converges in law to $T^{\rightarrow} := \sup_{p \in \Psi_0} \Upsilon(\mathcal{Z}(p,t))$, where $\Upsilon(x) = \{-x\}_+$ and $\Psi_0 = \mathbb{S}^{d-1}$. A critical value can be computed using Algorithm 3.1. Pointwise size control and the consistency against fixed alternatives then follow as a corollary to Theorem 3.2.

Corollary 3.1. *Suppose the conditions of Theorem 3.1 and Assumption 3.1 hold. Let Θ_0 be a nonempty compact convex subset of Θ^o. Let $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)$ be the $1-\alpha$ quantile of $\hat{F}_n(\cdot, t)$ computed by Algorithm 3.1 with $\Upsilon(x) = \{-x\}_+$ and $\Psi_n = \mathbb{S}^{d-1}$ for all n. Let $\tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t) + \delta$, where $\delta > 0$ is an arbitrarily small constant.*

(i) If $\Theta_0 \subseteq \Theta_I$ and $\alpha \in (0, 0.5)$, then it holds that

$\limsup_{n\to\infty} P\left(T_n^{\rightarrow}(t) > \tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right) \le \alpha$;

(ii) If $\Theta_0 \nsubseteq \Theta_I$, then the test is consistent: $\lim_{n\to\infty} P\left(T_n^{\rightarrow}(t) > \tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right) = 1$.

In Corollary 3.1 (and also in Corollary 3.2 below), an arbitrarily small constant $\delta > 0$ is introduced to the critical value. This is to ensure that the test remains asymptotically valid even if the limiting distribution $F^{\rightarrow}(\cdot, t)$ of the test statistic is not continuous at the $1-\alpha$ quantile or showing the continuity of $F^{\rightarrow}(\cdot, t)$ is not straightforward. However, if the limiting distribution is continuous, Corollary 3.1 holds with the critical value $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)$, and hence one does not need to introduce δ. In various empirically relevant examples, this is a reasonable assumption. For example, in the linear regression with an interval censored outcome (Example 1), the limiting distribution can be shown to be the maximum of a Gaussian process for which sufficient conditions for its absolute continuity are known (see e.g. Tsirel'son, 1976; Davydov, Lifshits, and Smorodina, 1998). Other examples that have the same structure include revealed preference bounds with linear payoffs (Pakes, 2010; Blundell, Kristensen, and Matzkin, 2014) and an IV model with a binary dependent variable (Chesher, 2009).

A one-sided confidence set $\mathcal{C}_n$ that covers the identified set with an asymptotic coverage probability $1-\alpha$ can be obtained by inverting the test in Corollary 3.1. For each t, define

$$\mathcal{C}_{1n}(t) := \{\theta \in \Theta : d(\theta, \hat{\Theta}_n(t)) \le \tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}\} \ , \qquad \tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t) + \delta \ . \tag{3.14}$$

This confidence set is an expansion of the set estimator $\hat{\Theta}_n(t)$ by the amount $\tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}$.[7] Under the conditions of Corollary 3.1, the confidence set satisfies $\liminf_{n\to\infty} P\left(\Theta_I \subseteq \mathcal{C}_{1n}(t)\right) \ge 1-\alpha$ for $\alpha \in (0, 0.5)$ and t small enough.

Remark 3.2. The hypothesis in (3.12) and its test using an estimated support function is first studied in BM. Here, we use the same hypothesis testing framework as theirs. However, there are two key differences. First, the class of models considered here is different. We require that Θ_I to be the set of minimizers of a convex criterion function, while BM considers a class of models in which the identified set can be represented as a linear function of the Aumann expectation of random sets. Second, the limiting distribution of the directed Hausdorff distance statistic $T_n^{\rightarrow}(t)$ is derived differently. We use Theorem 3.1 exploiting the duality between the support function and criterion function, while BM applies the central limit theorem for IID random sets and a continuous mapping theorem to their estimator, which is based on a sample average of IID random sets.

We outline below how to construct the confidence region in the context of Example 1.

Example 1 (Interval censored outcome (continued)). Constructing $\mathcal{C}_{1n}(t)$ requires the researcher to compute the support function of the consistent set estimator $\hat{\Theta}_n(t)$ and subsampled set estimators $\hat{\Theta}_{n,b,k}(t)$ in Algorithm 3.1. As pointed out earlier, these objects can be computed by solving convex programs. With the criterion function in (3.7), one can use the following linear program (LP) to compute support functions:

$$\begin{aligned} &\max_{(\theta,v)\in\mathbb{R}^d\times\mathbb{R}^{2K}} \langle p, \theta\rangle \\ &\text{subject to } \sum_{k=1}^{2K} w_k v_k \le t \ , \\ &\qquad\qquad - v_k + a_k'\theta \le b_k, k = 1, \cdots 2K \ , \\ &\qquad\qquad v_k \ge 0,\ k = 1, \cdots, 2K, \end{aligned} \tag{3.15}$$

where $b = (b_1, \cdots, b_{2K})$ determines the location of the linear constraints, $w = (w_1, \cdots, w_{2K})$ is a weight vector on the constraints, and $v = (v_1, \cdots, v_{2K})$ is a vector of auxiliary control variables.[8] For example, if we set $w_k = \sqrt{n}\hat{\sigma}_{k,n}^{-1}$ and $b_k = F_k(\hat{E}_n[m(X_i)])$ with $m(x) = (y_L 1_{\mathcal{Z}}(z), y_U 1_{\mathcal{Z}}(z), 1_{\mathcal{Z}}(z))$ for each k, the optimal value of the LP above yields $s(p, \hat{\Theta}_n(t))$, the support function of the set

[7] $\mathcal{C}_{1n}(t)$ can also be written as $\{\theta \in \Theta : \vec{d}_H(\{\theta\}, \hat{\Theta}_n(t)) \le \tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}\}$ because, when the first argument A of $\vec{d}_H$ is a singleton $\{a\}$, we have $\vec{d}_H(A,B) = \sup_{a\in A} d(a,B) = d(a,B)$ by Eq. (2.4).

[8] For more details on computation, see discussions in Appendix F.

estimator. Subsampled support functions $s(p, \hat{\Theta}_{n,b,k}(t))$ can be computed analogously. Hence, one may compute the normalized support function $\mathcal{Z}_{n,b,k}(\cdot, t)$ in Algorithm 3.1 and obtain a critical value $\tilde{c}_{n,b,1-\alpha}(t)$ as defined in Corollary 3.1. To compute the confidence region $\mathcal{C}_{1n}(t)$, solve the LP in (3.15) again while replacing the level t with $\tilde{c}_{n,b,1-\alpha}(t)$. This gives, for each $p \in \mathbb{S}^{d-1}$, a boundary point of the confidence region $\mathcal{C}_{1n}(t)$ as the optimizer of the problem. Hence, repeating this for different directions, one can trace out the boundary of $\mathcal{C}_{1n}(t)$.

The directed Hausdorff distance statistic based on $\hat{\Theta}_n(t)$ involves a user chosen parameter, the initial level t. As we will see in section 4.2, we can often properly weight the criterion function so that the level t only affects the mean of the limiting process $\mathcal{Z}(p, t)$. In this case, we can re-center the process $\mathcal{Z}_n(p, t)$ by a known function $\mu(t)$ or a consistent estimator $\hat{\mu}_n(t)$, so that the choice of level becomes asymptotically irrelevant for inference. Even if we do not have a known form for $\mu(t)$ or a consistent estimator, it is possible to remove the arbitrariness in the choice of t.

For each $\alpha \in (0, 1)$, let $c^{\rightarrow}_{1-\alpha}(t) = \inf\{x \in \mathbb{R} : P(T^{\rightarrow}(t) \leq x) \geq 1 - \alpha\}$ denote the $1 - \alpha$ quantile of the limit law $T^{\rightarrow}(t)$ of the test statistic in (3.13) and let

$$t^*_{1-\alpha} := \inf\{t \in \mathbb{R}_+ : c^{\rightarrow}_{1-\alpha}(t) = 0\}. \tag{3.16}$$

Lemma D.1 (in the appendix) shows $t \mapsto c^{\rightarrow}_{1-\alpha}(t)$ is non-increasing on the interval $[0, t^*_{1-\alpha}]$. This suggests that if we start with a large t, the amount used in (3.14) to expand the set estimator will be smaller asymptotically, and at $t = t^*_{1-\alpha}$, we do not need to expand the set at all. The following theorem gives conditions under which this change in the amount of expansion makes all confidence sets with $t \in [0, t^*_{1-\alpha})$ asymptotically equivalent. For this result, we require that the limiting distribution of $T^{\rightarrow}_n(t)$ is continuous.[9]

Theorem 3.3. *Suppose the conditions of Theorem 3.1 and Assumption 3.1 hold. Suppose that the limiting process takes the form $\mathcal{Z}(p, t) = \mu(t) + \mathcal{Z}^*(p)$ for each $(p, t) \in \mathbb{S}^{d-1} \times \mathbb{R}_+$ where $\mu : \mathbb{R}_+ \to \mathbb{R}$ is an unknown function and that $\mathcal{Z}_n(p, t) - \mathcal{Z}_n(p, t') = \mu(t) - \mu(t') + o_p(1)$ uniformly in p. Suppose that for each $t \in [0, t^*_{1-\alpha})$, the cdf of $T^{\rightarrow}(t)$ is continuous and strictly increasing at its $1-\alpha$ quantile. Let $\mathcal{C}_{1n}(t)$ be defined as in* (3.14) *where $\tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)$. Then (i) for each $\alpha \in (0, 1)$ and $0 \leq t < t^*_{1-\alpha}$,*

$$d_H\big(\mathcal{C}_{1n}(t), \hat{\Theta}_n(t^*_{1-\alpha})\big) = o_p(a_n^{-1/\gamma}). \tag{3.17}$$

(ii) for each $\alpha \in (0, 1)$ and for any $t, t' \in [0, t^*_{1-\alpha})$, it holds that $d_H\left(\mathcal{C}_{1n}(t), \mathcal{C}_{1n}(t')\right)) = o_p(a_n^{-1/\gamma})$.

We also propose a generic iterative algorithm to construct a confidence set.

Algorithm 3.2. (Iterative Algorithm) Set $\kappa > 0$ small. Initialize $l = 1$, and set t_l to an initial value, say $t_l = 0$.

[9] Recall that, if the limiting distribution of $T^{\rightarrow}_n(t)$ is continuous, the conclusions of Corollary 3.1 (i) holds with the critical value $\tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)$, i.e. $\delta = 0$ in (3.14).

Step 1. Construct the set estimator $\hat{\Theta}_n(t_l)$. Estimate the asymptotic $1-\alpha$ quantile $c^{\rightarrow}_{1-\alpha}(t_l)$ of the scaled directed Hausdorff distance $a_n^{1/\gamma}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t_l))$ by Algorithm 3.1 with $\Upsilon(x) = \{-x\}_+$ and $\Psi_0 = \mathbb{S}^{d-1}$, obtaining $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t_1)$. Using $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t_l)$, expand $\hat{\Theta}_n(t_l)$ by $\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t_l) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t_l)/a_n^{1/\gamma}$ to obtain $\mathcal{C}_{1n}(t_l)$.

Step 2. Update the level by setting $t_{l+1} := \sup_{\theta\in\mathcal{C}_{1n}(t_l)} a_n Q_n(\theta)$.

Step 3. Repeat steps 1-2 until $|t_{l+1} - t_l| < \kappa$.

The iterative algorithm can be proved to yield an increasing sequence $\{t_l, l = 1, 2, \cdots\}$ that tends to $t^*_{1-\alpha}$. As Theorem 3.3 shows, if the limiting process takes the form $\mathcal{Z}(p,t) = \mu(t) + \mathcal{Z}^*(p)$, one may stop at Step 1.

Remark 3.3. CHT's confidence set satisfies $\lim_{n\to\infty} P(\Theta_I \subseteq \hat{\Theta}_n(\hat{\tau}_{n,b,1-\alpha})) = 1-\alpha$, where $\hat{\tau}_{n,b,1-\alpha}$ is a subsampling estimate of the $1-\alpha$ quantile $\tau^*_{1-\alpha}$ of the limiting distribution of their QLR-statistic $S_n := \sup_{\Theta_I} a_n Q_n(\theta)$. Theorem 3.3 suggests, if $t^*_{1-\alpha} = \tau^*_{1-\alpha}$, the confidence sets based on the QLR-approach and our approach are asymptotically equivalent. In section 4.1, we will provide conditions under which this holds for moment inequality models.

3.2.2 Inference for Points in the Identified Set

The estimated support function can also be used to make inference for points in the identified set. Let $\theta_0 \in \Theta$, and consider testing

$$H_0 : \theta_0 \in \Theta_I \quad vs. \quad H_1 : \theta_0 \notin \Theta_I. \tag{3.18}$$

We again use the directed Hausdorff distance statistic to test the hypothesis. Define the statistic

$$T^{\rightarrow}_{n,\theta_0}(t) := a_n^{1/\gamma}\vec{d}_H(\{\theta_0\}, \hat{\Theta}_n(t)) = \sup_{p\in\mathbb{S}^{d-1}} a_n^{1/\gamma}\big\{\langle p, \theta_0\rangle - s(p, \hat{\Theta}_n(t))\big\}_+. \tag{3.19}$$

The following theorem characterizes the asymptotic distribution of this statistic when θ_0 is on the boundary of Θ_I.

Theorem 3.4. *Suppose the conditions of Theorem 3.1 hold. Suppose further that $\theta_0 \in \partial\Theta_I$. Then,*

$$T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{d} \sup_{p\in\Psi_0}\{-\mathcal{Z}(p,t)\}_+, \tag{3.20}$$

where $\Psi_0 \subseteq \mathbb{S}^{d-1}$ is defined as $\Psi_0 := argmax_{p\in\mathbb{S}^{d-1}}\langle p, \theta_0\rangle - s(p, \Theta_I)$.

Let $c^{\rightarrow}_{1-\alpha}(\theta_0, t)$ be the $1-\alpha$ quantile of $\sup_{p\in\Psi_0}\{-\mathcal{Z}(p,t)\}_+$. An aspect specific to pointwise inference is that Ψ_0 in (3.20) is generally unknown and hence needs to be estimated from data. This is, however, straightforward. Since Ψ_0 is the set of maximizers of a criterion function, it admits consistent estimation by a level-set estimator. Letting $\{\kappa_n\}$ be a sequence of positive constants such that $\kappa_n \to \infty$ and $\kappa_n / a_n^{1/\gamma} \to 0$, we define

$$\hat{\Psi}_n := \{p \in \mathbb{S}^{d-1} : \langle p, \theta_0 \rangle - s(p, \hat{\Theta}_n(t)) \leq \sup_{p'}(\langle p', \theta_0 \rangle - s(p', \hat{\Theta}_n(t))) - \kappa_n / a_n^{1/\gamma}\}. \tag{3.21}$$

The following theorem establishes that the test has asymptotic level α and is consistent against any fixed alternative hypothesis.

Corollary 3.2. *Suppose the conditions of Theorem 3.2 hold. Let* $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(\theta_0, t)$ *be the* $1-\alpha$ *quantile of* $\hat{F}_n(\cdot, t)$ *computed by Algorithm 3.1 with* $\Upsilon(x) = \{-x\}_+$ *and* $\Psi_n = \hat{\Psi}_n$ *for all* n*. Let* $\tilde{c}^{\rightarrow}_{n,b,1-\alpha}(\theta_0, t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\theta_0, t) + \delta$*, where* $\delta > 0$ *is an arbitrarily small constant.*

(i) If $\theta_0 \in \Theta_I$ *and* $\alpha \in (0, 0.5)$*, then it holds that*

$\limsup_{n\to\infty} P\left(T^{\rightarrow}_{n,\theta_0}(t) > \tilde{c}^{\rightarrow}_{n,b,1-\alpha}(\theta_0, t)\right) \leq \alpha$;

(ii) If $\theta_0 \notin \Theta_I$*, then for any* $t \in \mathbb{R}_+$ *and* $\alpha \in (0, 1)$*, the test is consistent*:

$\lim_{n\to\infty} P\left(T^{\rightarrow}_{n,\theta_0}(t) > \tilde{c}^{\rightarrow}_{n,b,1-\alpha}(\theta_0, t)\right) = 1.$

A confidence set for θ_0 can be obtained by inverting the test in Corollary 3.2. Define

$$\mathcal{C}_{2n}(t) := \{\theta \in \Theta : T^{\rightarrow}_{n,\theta}(t) \leq \tilde{c}^{\rightarrow}_{n,b,1-\alpha}(\theta, t)\}. \tag{3.22}$$

Under the conditions of Theorem 3.2, this confidence set has the coverage property:

$$\liminf_{n\to\infty} P\left(\theta_0 \in \mathcal{C}_{2n}(t)\right) \geq 1 - \alpha, \quad \text{for all } \theta_0 \in \Theta_I. \tag{3.23}$$

Remark 3.4. A statistic closely related to $T^{\rightarrow}_{n,\theta_0}(t)$ is studied in Bontemps, Magnac, and Maurin (2012, Proposition 10) in the context of the incomplete linear model. To derive the asymptotic distribution of their statistic, these authors construct a sequence p_n of unit vectors that converges to some $p_0 \in \Psi_0$. Theorem 3.4 is a novel result that complements their work by deriving the asymptotic distribution of the statistic without such a sequence. For this, we note that our statistic can be written as

$$T^{\rightarrow}_{n,\theta_0}(t) = \max\{a_n^{1/\gamma}(\phi_{\theta_0}(s(p, \hat{\Theta}_n(t))) - \phi_{\theta_0}(s(p, \Theta_I))), 0\},$$

where for any $x : \mathbb{S}^{d-1} \to \mathbb{R}$, $\phi_{\theta_0}(x) := \sup_{p \in \mathbb{S}^{d-1}} \langle p, \theta_0 \rangle - x(p)$. In the appendix, we show that ϕ_{θ_0} belongs to a class of Hadamard directionally differentiable functionals. Theorem 3.1 and a functional δ-method in Shapiro (1991) then imply Theorem 3.4.

4 Moment Inequality Models

4.1 Inference for Moment Inequality Models

We apply the main results to models defined by finitely many moment inequalities. This class has been extensively studied recently.[10] We first show that, employing a criterion function used in CHT, the framework in Section 3 can be applied to convex moment inequalities. We then characterize the limiting distribution of the normalized support function. This ensures that the researcher may apply the Wald inference methods developed in the previous section to this class of models. In Section 4.2, we further establish a close connection between the support function and criterion function approaches using the characterization of the limiting distribution. Specifically, we show that inference based on Wald and QLR statistics become asymptotically equivalent under some conditions. This result can be thought of as a generalization of an asymptotic equivalence result Beresteanu and Molinari (2008) established for interval identified models, a special case of convex moment inequalities.

In the following, we use E and $\hat{E}_n$ to denote the expectation operators with respect to the data generating probability measure and the empirical measure, respectively. Let $m_j : \mathbb{R}^k \times \mathbb{R}^d \to \bar{\mathbb{R}}$, $j = 1, \cdots, J$ and m_θ be a $J \times 1$ vector whose j-th component is $m_{j,\theta} := m_j(X;\theta)$. The model is then characterized by the following moment inequality restrictions:

$$E(m_j(X;\theta)) \leq 0, \quad j = 1, \cdots, J. \tag{4.1}$$

A number of examples including Example 2 have this structure. The identified set Θ_I is the set of parameter values at which these restrictions are satisfied. Following CHT, we consider population and sample criterion functions of the form:

$$Q(\theta) = \|W^{1/2}(\theta)E(m_\theta)\|_+^2, \quad \text{and} \quad Q_n(\theta) = \|\hat{W}_n^{1/2}(\theta)\hat{E}_n(m_\theta)\|_+^2, \tag{4.2}$$

where W and $\hat{W}_n$ are population and sample weighting matrices. Below, we let $\bar{\mathcal{P}}_J$ be the set of $J \times J$ positive definite matrices and make the following assumptions (Assumptions 4.1-4.3), which are based on Condition M.2 in CHT. These conditions ensure the high-level conditions (Assumptions 2.1-2.3).

Assumption 4.1. *Let $(\Omega, \mathfrak{F}, P)$ be a complete probability space. Let $d \in \mathbb{N}$, and let $\Theta \subseteq \mathbb{R}^d$ be compact and convex, with a nonempty interior; (ii) $\theta \mapsto E(m_\theta)$ is continuous. $W : \mathbb{R}^d \to \bar{\mathcal{P}}_J$ is finite and continuous on Θ, and* $\det(W(\theta)) = \infty$ *if $\theta \notin \Theta$; (iii) $\hat{W}_n : \Omega \times \mathbb{R}^d \to \bar{\mathcal{P}}_J$ is finite and continuous on Θ, uniformly in n, and* $\det \hat{W}_n(\omega, \theta) = \infty$ *if $\theta \notin \Theta$ with probability 1.*

Assumption 4.1 makes a continuity assumption on the population moment function $\theta \mapsto E(m_\theta)$

[10] Recent research in this area includes Guggenberger, Hahn, and Kim (2008), Rosen (2008), Andrews and Guggenberger (2009), Galichon and Henry (2009), Canay (2010), Bugni (2010), Andrews and Soares (2010), Fan and Park (2010), Pakes, Porter, Ho, and Ishii (2011), Andrews and Barwick (2012), Moon and Schorfheide (2012), and Yildiz (2012) among others.

and mild regularity conditions on the parameter space and weighting matrix so that the population criterion function is well-defined. The next assumption then assumes that $\{m_\theta\}$ is a P-Donsker class and a uniform consistent estimator of the weighting matrix is available. These conditions can be satisfied by various moment functions (van der Vaart and Wellner, 2000) and estimators of weighting matrices, e.g. $\hat{W}_n(\theta)$ is an estimator of the inverse of the asymptotic covariance matrix of $\hat{E}_n[m_\theta]$.

Assumption 4.2. *(i) $\{m_\theta : \theta \in \Theta\}$ is a P-Donsker class; (ii) $\hat{W}_n(\theta) - W(\theta) = o_p(1)$ uniformly over Θ.*

We further assume that the population moment function decreases in the interior of the identified set and increases outside a neighborhood of the identified set sufficiently rapidly. This assumption is used to guarantee the existence of a polynomial minorant in Assumption 2.2 and a sequence of sets Θ_n in Assumption 2.3, on which the sample criterion function degenerates.

Assumption 4.3. *(i) There exist positive constants $(C, M, \bar{\epsilon})$ such that for any $0 \le \epsilon \le \bar{\epsilon}$ and $\theta \in \Theta_I^{-\epsilon}$, $\max_{1\le j\le J} E(m_{j,\theta}) \le -C\epsilon$, and $d_H(\Theta_I^{-\epsilon}, \Theta_I) \le M\epsilon$, where $\Theta_I^{-\epsilon} = \{\theta \in \Theta_I : d(\theta, \Theta \setminus \Theta_I) \ge \epsilon\}$; (ii) There exist positive constants (C, δ) such that for any $\theta \in \Theta$, $\|E(m_\theta)\|_+ \ge C(d(\theta, \Theta_I) \wedge \delta)$, and a continuous Jacobian $\Pi(\theta) := \nabla_\theta E[m_\theta]$ exists for each $\theta \in \Theta$.*

Finally, we assume that the identified set is in the interior of Θ and that the population and sample criterion functions are convex.

Assumption 4.4. *(i) The map $\theta \mapsto \|W^{1/2}(\theta)E(m_\theta)\|_+^2$ is convex; (ii) $\theta \mapsto \|\hat{W}_n^{1/2}(\theta)\hat{E}_n(m_\theta)\|_+^2$ is convex a.s.; (iii) $\Theta_I \subset \Theta^o$.*

Under these assumptions, the localized criterion function based on Q_n in (4.2) can be written as:

$$\begin{aligned}
\ell_n(\theta,\lambda) &= \left\|\hat{W}_n^{1/2}(\theta+\lambda/\sqrt{n})\sqrt{n}\hat{E}_n(m_{\theta+\lambda/\sqrt{n}})\right\|_+^2 \\
&= \left\|\hat{W}_n^{1/2}(\theta+\lambda/\sqrt{n})(\mathbb{G}_n m_{\theta+\lambda/\sqrt{n}} + E[m_{\theta+\lambda/\sqrt{n}}])\right\|_+^2 = \left\|\hat{W}_n^{1/2}(\theta+\lambda/\sqrt{n})\mathcal{M}_n(\theta,\lambda)\right\|_+^2, \quad (4.3)
\end{aligned}$$

where $\mathcal{M}_n(\theta,\lambda) := \mathbb{G}_n m_{\theta+\lambda/\sqrt{n}} + \Pi(\bar{\theta}_n)\lambda + \sqrt{n}E(m_\theta)$, $\mathbb{G}_n m_\theta := \sqrt{n}(\hat{E}_n[m_\theta] - E[m_\theta])$ is an empirical process indexed by $\theta \in \Theta$, and $\bar{\theta}_n$ is a mean value which lies between θ and $\theta + \lambda/\sqrt{n}$. One may then show that ℓ_n converges to the following limit in the mode required by Theorem 3.1:

$$\ell_\infty(\theta,\lambda) = \|W^{1/2}(\theta)\mathcal{M}(\theta,\lambda)\|_+^2, \quad (\theta,\lambda) \in \partial\Theta_I \times \mathbb{R}^d\ , \tag{4.4}$$

where $\mathcal{M}(\theta,\lambda) = \mathbb{G}(\theta) + \Pi(\theta)\lambda + \varsigma(\theta)$, $\mathbb{G}(\cdot)$ is a Gaussian process on Θ, and $\varsigma(\theta)$ is a vector whose j-th component is such that, for any $\theta \in \partial\Theta_I$, $\varsigma_j(\theta) = -\infty$ when the population constraint is slack, i.e. $E(m_{j,\theta}) < 0$, and $\varsigma_j(\theta) = 0$ when the population constraint binds, i.e. $E(m_{j,\theta}) = 0$. The following theorem shows that our previously stated high-level conditions are satisfied under Assumptions 4.1-4.4.

Theorem 4.1. *Suppose Assumptions 4.1-4.4 hold. Then Assumptions 2.1-2.4 and B.1 are satisfied with ℓ_∞ in* (4.4).

Theorem 3.1 then applies. Therefore, the normalized support function $\mathcal{Z}_n(\cdot, t)$ converges in law to the following process:

$$\mathcal{Z}(p,t) = \sup_{\theta \in H(p,\Theta_I)} \sup_{\lambda \in \{\lambda : \|W^{1/2}(\theta)\mathcal{M}(\theta,\lambda)\|_+^2 \le t\}} \langle p, \lambda \rangle. \tag{4.5}$$

Hence, we have obtained a characterization of the limiting distribution for the normalized support function. This has two implications. First, the existence of the limiting distribution allows one to employ the Wald-type methods developed in Section 3 to make inference within models defined by convex moment inequalities. Second, (4.5) is also useful for establishing a further connection between the support function and criterion function approaches, which we elaborate below.

4.2 Asymptotic Equivalence of Wald and QLR Statistics

We use (4.5) to show that, under additional assumptions, the Wald statistic (squared directed Hausdorff distance) and CHT's QLR statistic are asymptotically equivalent. Although this requires additional restrictions, they allow us to establish a conceptually important connection between the two approaches. Toward this end, we introduce additional notation to denote active and slack moment inequalities. For each $\theta \in \partial\Theta_I$, let $\mathcal{J}(\theta) := \{j \in \{1, \cdots, J\} : E(m_{j,\theta}) = 0\}$ be the set of indices associated with active moment inequalities, and let $J(\theta)$ be the number of elements in $\mathcal{J}(\theta)$. Let $\Pi_{\mathcal{J}(\theta)}(\theta)$ denote the $J(\theta) \times d$ matrix that stacks rows of $\Pi(\theta)$ whose indices belong to $\mathcal{J}(\theta)$. Let $\mathbb{G}_{\mathcal{J}(\theta)}$ denote the $J(\theta) \times 1$ vector of Gaussian processes that stacks components of $\mathbb{G}$ whose indices belong to $\mathcal{J}(\theta)$. Finally, let $W_{\mathcal{J}(\theta)}$ denote the $J(\theta) \times J(\theta)$ matrix that collects (i,j) elements of $W(\theta)$ for $i, j \in \mathcal{J}(\theta)$.

In the current setting, the support function $s(\cdot, \Theta_I)$ of the identified set is the optimal value function of the following problem:

$$\begin{aligned} \sup \quad & \langle p, \theta \rangle \\ s.t. \quad & E(m_{j,\theta}) \le 0, \quad \text{for } j = 1, \cdots, J. \end{aligned} \tag{4.6}$$

Eq. (4.5) implies that the limiting distribution of $\mathcal{Z}_n(\cdot, t)$ can be studied by analyzing the following approximating problem for each $p \in \mathbb{S}^{d-1}$ and $\theta \in H(p, \Theta_I)$:

$$\begin{aligned} \sup_{\lambda} \quad & \langle p, \lambda \rangle \\ s.t. \quad & \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)[\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda]\|_+^2 \le t. \end{aligned} \tag{4.7}$$

Here, we note that the only binding constraints, i.e. $j \in \mathcal{J}(\theta)$, are relevant in (4.7). This is because the slack inequalities are dropped because $\varsigma_j(\theta) = -\infty$ for any $j \notin \mathcal{J}(\theta)$, and the criterion function discards negative moments. Solving the optimization problem in (4.7) then gives a closed form for

$\mathcal{Z}(p,t)$. For this, we add the following assumption to simplify the limiting distribution.

Assumption 4.5. *For each $\theta \in \partial\Theta_I$, $rank(\Pi_{\mathcal{J}(\theta)}) = J(\theta)$, i.e. the rows of the Jacobian matrices are linearly independent.*

In Assumption 4.5, we assume that the gradients of the binding moment inequalities are linearly independent at each boundary point. This, for example, excludes the case where some boundary point is formed by the intersection of more than d inequalities. Using this assumption and (4.7), one can then simplify the representation of the limiting distribution, which allows to compare the weak limit of the Wald statistic with that of CHT's QLR statistic: $\sup_{\Theta_I} a_n Q_n(\theta)$.

Corollary 4.1 (Asymptotic Equivalence for Moment Inequalities)**.** *Suppose Assumptions 4.1-4.5 and E.1 in the appendix hold. Suppose $W(\theta)$ satisfies $W_{\mathcal{J}(\theta)}(\theta) = [\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)']^{-1}$ for each $\theta \in \partial\Theta_I$. Suppose Θ_I is strictly convex. For each $p \in \mathbb{S}^{d-1}$, let $\theta_I(p) \in \partial\Theta_I$ be such that $H(p,\Theta_I) = \{\theta_I(p)\}$.*

Then, (i) $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}_n(p,t) + t^{1/2}\}_+^2 \xrightarrow{d} \mathbf{Z}$ *and* $\sup_{\Theta_I} nQ_n(\theta) \xrightarrow{d} \mathbf{Z}$, *where*

$$\mathbf{Z} := \sup_{p\in\mathbb{S}^{d-1}} \left\langle \left(\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))'\right)^{-1}\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))p, \mathbb{G}_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\right\rangle_+^2 ;$$

(ii) Further, it holds that $t^*_{1-\alpha} = \tau^*_{1-\alpha}$.

Corollary 4.1 shows that the Wald statistic (squared directed Hausdorff distance) and CHT's QLR statistic are asymptotically equivalent in the sense that they converge in distribution to the same limit, a continuous functional of a Gaussian process. The second result implies the asymptotic equivalence of the Wald and QLR confidence sets for Θ_I. When $t^*_{1-\alpha} = \tau^*_{1-\alpha}$ for $\alpha \in (0,1/2)$, Theorem 3.3 implies that for all $t \in [0,\tau^*_{1-\alpha})$,[11]

$$d_H\left(\mathcal{C}_{1n}(t), \hat{\Theta}_n(\tau^*_{1-\alpha})\right) = o_p(n^{-1/2}) \ .$$

This means that the Wald confidence set, which is an expansion of the set estimator is asymptotically equivalent to the QLR confidence set, a level set using an asymptotic quantile of the QLR statistic as a level.

Remark 4.1. Corollary 4.1 can be viewed as a generalization, to the convex moment inequality models, of Theorem 3.1 in BM who establish an asymptotic equivalence of Wald and QLR statistics for models in which the identified set for a scalar parameter θ is defined by two moment inequalities $E(X_1) \leq \theta \leq E(X_2)$. BM use an estimator $\tilde{\Theta}_n$ based on the average of set-valued random variables, which is not necessarily a level-set estimator in general. However, in this special case, BM's estimator can be shown to coincide with the set of minimizers of the criterion function $Q_n(\theta) = (\hat{E}_n[X_{1i}] - \theta)_+^2 + (\theta - \hat{E}_n[X_{2i}])_+^2$; this therefore becomes a level-set estimator with $t = 0$. Hence, the "exact" equivalence of our Wald statistic $\mathcal{W}_n := \sqrt{n}\vec{d}_H(\Theta_I, \hat{\Theta}_n(0))$ and BM's Wald

[11] Theorem 3.3 is applicable with $\tilde{c}_{n,b,k}(t) = \hat{c}_{n,b,k}(t)$. This is because, for $\alpha \in (0,1/2)$, the continuity of the limiting distribution $\mathbf{Z}$ follows from Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998).

statistic $\tilde{\mathcal{W}}_n := \sqrt{n}\vec{d}_H(\Theta_I, \tilde{\Theta}_n)$ holds. Corollary 4.1 then implies that our Wald statistic, CHT's QLR statistic and BM's Wald statistic are all asymptotically equivalent in this special case.

5 Monte Carlo Experiments

We use Example 1 to examine the performance of our inference procedure. Let $Z_i \equiv (Z_{1,i}, Z_{2,i})'$ where $Z_{1,i} = 1$ and $Z_{2,i}$ is uniformly distributed on a set of K equally spaced points on $[-5, 5]$. For $\theta_0 = (1, 2)'$ we generate:

$$Y_i = Z_i'\theta_0 + \epsilon_i \qquad i = 1, \dots, n \ , \tag{5.1}$$

where ϵ_i is a standard normal random variable independent of Z_i. We then create upper and lower bounds $(Y_{L,i}, Y_{U,i})$ such that $Y_{L,i} \leq Y_i \leq Y_{U,i}$ by:

$$\begin{aligned} Y_{L,i} &= Y_i - C - V_i Z_i^2 \qquad i = 1, \dots, n \\ Y_{U,i} &= Y_i + C + V_i Z_i^2 \qquad i = 1, \dots, n \ , \end{aligned} \tag{5.2}$$

where $C = 1$ and V_i is uniformly distributed on $[0, 0.2]$ independently of (Y_i, Z_i). Using the notation introduced in Section 2.3, one may then define a sample criterion function as follows (see also (3.7) and subsequent discussions):

$$Q_n(\theta) = \sum_{k=1}^{2K} \hat{\sigma}_{k,n}^{-1} \Big(a_k'\theta - F_k(\hat{E}_n[m(X_i)] \Big)_+ , \tag{5.3}$$

For $x \in \mathbb{R}^{2K}$, we use the criterion function $\sum_{k=1}^{2K}(x_k)_+$ to aggregate binding moments instead of $\|x\|_+^2$ used in Section 4.1. This is to make a comparison to the bootstrap procedure by Bugni (2010) who uses this criterion function. Given a level set estimator $\hat{\Theta}_n(t)$ of the sample criterion function in (3.7), the Wald-statistic is defined by $T_n^{\rightarrow} := \sup_{p \in \mathbb{S}^{d-1}} \{\sqrt{n}(s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I))\}_+$. The Wald approach obtains a confidence set by expanding the set estimator $\hat{\Theta}_n(t)$ by the amount $\hat{c}_{n,b,1-\alpha}(t)/\sqrt{n}$ computed by subsampling the Wald statistic. The QLR statistic is defined by $S_n = \sup_{\Theta_I} Q_n(\theta)$. The QLR approach constructs a confidence set by taking a level set $\hat{\Theta}_n(\hat{\tau}_{n,b,1-\alpha})$, where the right level $\hat{\tau}_{n,b,1-\alpha}$ is computed by resampling the QLR statistic (CHT, 2007, Bugni, 2009).

We report the coverage probabilities of the following four confidence sets. The first confidence set $\mathcal{C}_{\text{Wald}}$ is a Wald-type confidence set defined as in (3.14) with $t = \ln(\ln(n))^{\frac{1}{2}}$ and $\delta = 0$. The second confidence set $\mathcal{C}_{\text{Iter}}$ is defined in the same manner but uses Algorithm 3.2 to update the initial level. $\mathcal{C}_{\text{CHT-Sub}}$ is CHT's confidence set with a subsampling critical value, and $\mathcal{C}_{\text{CHT-Boot}}$ is also a CHT-type confidence set with a critical value computed by a bootstrap procedure proposed by Bugni (2010). Details on the implementation of these procedures are discussed in Appendix F.

Table 1 and 2 report the results of the Monte Carlo experiments. Table 1 shows the coverage probabilities of the four confidence sets, and Table 2 reports the median of the Hausdorff distances between the identified set and each of the four confidence sets. For the first three confidence sets, we report the results under three different values of subsamples: (e.g. $b = 100, 150, 200$ for $n = 1,000$).

For the last confidence set $\mathcal{C}_{\text{CHT-Boot}}$, the procedure requires a tuning parameter κ_n, which is used to select moments that are relevant for the calculation of the critical value. We set this parameter to three different values: $\ln(\ln(n))^{\frac{1}{2}}, \ln(n)^{\frac{1}{2}}$, and $n^{1/8}$.

For $n = 1,000$, the coverage probability of Wald confidence set $\mathcal{C}_{\text{Wald}}$ tends to be slightly under the nominal level 0.95, but the size distortion is limited (0.1-4%) across all values of K. The size distortion, however, tends to become larger (2.1-5.3%) with a smaller sample size: $n = 500$. The coverage probability of $\mathcal{C}_{\text{Iter}}$ is higher than that of $\mathcal{C}_{\text{Wald}}$ and has a better size control property although it is fairly conservative under some choice of subsample sizes. This can also be seen from Table 2, which shows that the median Hausdorff loss of $\mathcal{C}_{\text{Iter}}$ is larger than that of $\mathcal{C}_{\text{Wald}}$. One thing to note is that when the number of moment inequalities is large (30 inequalities), $\mathcal{C}_{\text{Wald}}$ tends to be quite conservative for some subsample sizes (e.g. $b = 100, 150$ with $n = 500$). This indicates that the asymptotic approximation with a subsampling critical value may not provide a good approximation to the finite sample distribution when the number of inequalities is large.

The coverage probabilities of the CHT confidence sets vary with subsample sizes. $\mathcal{C}_{\text{CHT-Sub}}$ has size distortion in some settings (e.g. $K = 5$). As pointed out by Bugni (2009), a better size control is achieved by using a suitable bootstrap procedure. The coverage probability of $\mathcal{C}_{\text{CHT-Boot}}$ controls the size across all values of K, which therefore shows the uniform validity of the procedure. The result is also robust over the set of values of tuning parameters, while in some cases the procedure is conservative. Overall, the size of Wald confidence set $\mathcal{C}_{\text{Wald}}$ is not as good as the CHT's confidence set with a suitable bootstrap procedure. The confidence set $\mathcal{C}_{\text{Iter}}$ with the iterative algorithm has a better size control property than $\mathcal{C}_{\text{Wald}}$, but it is conservative in some settings.

6 Conclusion

This paper introduces a framework for partially identified econometric models that unifies two general approaches recently proposed in the literature: the criterion function approach and the support function approach. We consider the general case where the convex identified set Θ_I is the set of minimizers of a criterion function, estimated as an appropriate level set of a sample criterion function, following CHT, and represented as a support function, as in BM. Our main duality result shows that the support function of CHT's level set estimator converges to a well-defined limit when a localized criterion function converges in a suitable manner. This yields Wald-type inference methods based on the estimated support function for general models whose identified sets can be characterized as the set of minimizers of convex criterion functions, which therefore allow to study examples that do not belong to the models studied by BM and Bontemps, Magnac, and Maurin (2012).

We highlight the duality result within the class of moment inequality models by establishing the asymptotic equivalence of our Wald statistic and CHT's QLR statistic. We further show that this implies the asymptotic equivalence of the Wald confidence set and CHT's confidence set. For inference on the identified set and points inside it, we propose a general subsampling procedure. This procedure is valid pointwise, as we derive our results under a fixed probability measure. Establishing the uniform asymptotic validity of subsampling is important for partially identified models and is one of our future tasks.

References

Aliprantis, C. D., and K. C. Border (2006): *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

Andrews, D. W. K., and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80(6), 2805–2826.

Andrews, D. W. K., and P. Guggenberger (2009): "Validity of Subsampling and 'Plug-In Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25(3), 669–709.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

Blundell, R., D. Kristensen, and R. L. Matzkin (2014): "Bounding quantile demand functions using revealed preference inequalities," *Journal of Econometrics*, 179(2), 112–127.

Bontemps, C., T. Magnac, and E. Maurin (2012): "Set Identified Linear Models," *Econometrica*, 80(3), 1129–1155.

Boyd, S., and L. Vandenberghe (2004): *Convex Optimization*. Cambridge University Press, Cambridge.

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

Canay, I. (2010): "EL inference for partially identified models: large deviations optimality and bootstrap validity," *Journal of Econometrics*, 156(2), 408–425.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Chesher, A. (2009): "Single Equation Endogenous Binary Response Models," Discussion paper, Cemmap working paper.

Davydov, I. A., M. A. Lifshits, and N. V. Smorodina (1998): *Local Properties of Distributions of Stochastic Functionals (Translations of Mathematical Monographs)*. American Mathematical Society.

Fan, Y., and S. S. Park (2010): "Confidence Sets for Some Partially Identified Parameters," Discussion Paper, Vanderbilt University.

Galichon, A., and M. Henry (2009): "A Test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics*, 152(2), 186 – 196.

Guggenberger, P., J. Hahn, and K. Kim (2008): "Specification testing under moment inequalities," *Economics Letters*, 99(2), 375–378.

Kaido, H., and A. Santos (2014): "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities," *Econometrica*, 82(1), 387–413.

Kaido, H., and H. White (2009): "Inference on Risk Neutral Measures for Incomplete Markets," *Journal of Financial Econometrics*, 7(3), 199–246.

Knight, K. (1999): "Epi-convergence in Distribution and Stochastic Equi-semicontinuity," Discussion Paper, University of Tronto.

Lehmann, E., and J. P. Romano (2005): *Testing Statistical Hypotheses (Springer Texts in Statistics)*. Springer-Verlag, Berlin, 3rd edn.

Li, S., Y. Ogura, and V. Kreinovich (2002): *Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables (Theory and Decision Library B)*. Kluwer Academic Publishers, Dordrecht.

Luttmer, E. (1996): "Asset Pricing in Economies with Frictions," *Econometrica*, 64(6), 1439–1467.

Magnac, T., and E. Maurin (2008): "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *Review of Economic Studies*, 75(3), 835–864.

Manski, C. F. (2003): *Partial Identification of Probability Distributions (Springer Series in Statistics)*. Springer-Verlag, Berlin.

Manski, C. F., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

Molchanov, I. (2005): *Theory of Random Sets*. Springer-Verlag, Berlin.

Moon, H. R., and F. Schorfheide (2012): "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80(2), 755–782.

Newey, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59(4), 1161–1167.

Pakes, A. (2010): "Alternative Models for Moment Inequalities," *Econometrica*, 78(6).

Pakes, A., J. Porter, K. Ho, and J. Ishii (2011): "Moment Inequalities and Their Application," Discussion Paper, Harvard University.

Politis, D., J. Romano, and M. Wolf (1999): *Subsampling*. Springer-Verlag, Berlin.

Rockafellar, R. T. (1970): *Convex Analysis*. Princeton University Press, Princeton.

Rockafellar, R. T., and R. J.-B. Wets (2005): *Variational Analysis, Second Edition*. Springer-Verlag, Berlin.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146(1), 107 – 117.

Shapiro, A. (1991): "Asymptotic analysis of stochastic programs," *Annals of Operations Research*, 30(1), 169–186.

Tsirel'son, V. (1976): "The density of the distribution of the maximum of a Gaussian process," *Theory of Probability & Its Applications*, 20(4), 847–856.

van der Vaart, A., and J. Wellner (2000): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, Berlin.

Yildiz, N. (2012): "Consistency of Plug-In Estimators of Upper Contour and Level Sets," *Econometric Theory*, 28(02), 309–327.

Zalinescu, C. (2002): *Convex Analysis in General Vector Spaces*. World Scientific Publishing Company, Singapore.

Supplemental Appendix

In this supplemental appendix, we include additional conditions and the proofs of the results stated in the main text. The contents of the supplemental appendix are organized as follows. Appendix A collects notation and definitions used throughout the appendix. Appendix B gives the local process regularity, and Appendix C contains the proof of Theorem 3.1 and auxiliary lemmas. Appendix D contains the proof of Theorems 3.2, 3.3, 3.4, Corollary 3.2, and auxiliary lemmas. Appendix E contains the proof of Theorem 4.1, Corollaries E.1 and 4.1. Appendix F collects details on the Monte Carlo experiments.

Appendix A: Notation and Definitions

The following is a list of notations and definitions used throughout the appendix.

ℓ_n :	Localized criterion function defined as $\ell_n(\theta,\lambda) = a_n Q_n(\theta + \lambda/a_n^{1/\gamma})$.
ℓ_∞ :	Limit of ℓ_n in the mode of Assumption B.1 (ii).
$\Lambda_{\theta,t}$:	The level set $\Lambda_{\theta,t} = \{\lambda \in \mathbb{R}^d : \ell_\infty(\theta,\lambda) \le t\}$.
$H(p,\Theta_I)$:	The support set $H(p,\Theta_I) = \{\theta : \langle p,\theta\rangle = s(p,\Theta_I)\} \cap \Theta_I$.
$K_{u,p}, K^o_{u,p}$:	The half spaces $K_{u,p} := \{x \in \mathbb{R}^d : \langle p,x\rangle \ge u\}$ and $K^o_{u,p} = \{x \in \mathbb{R}^d : \langle p,x\rangle > u\}$.
$R_{u,p}, R^o_{u,p}$:	The sets $R_{u,p} = H(p,\Theta_I) \times K_{u,p}$ and $R^o_{u,p} = H(p,\Theta_I) \times K^o_{u,p}$.
$R_{n,u,p}, R^o_{n,u,p}$:	The sets $R_{n,u,p} = \{(\theta,\lambda) : \theta \in H(p,\Theta_I), \lambda \in K_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta)\}$ and $R^o_{n,u,p} = \{(\theta,\lambda) : \theta \in H(p,\Theta_I), \lambda \in K^o_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta)\}$.
$\mathcal{J}(\theta)$:	The set of indices associated with binding moment inequalities at $\theta \in \Theta$.
$J(\theta)$:	The number of elements in $\mathcal{J}(\theta)$.
$\Pi_{\mathcal{J}(\theta)}$:	The $J(\theta) \times d$ matrix that stacks rows of $\Pi(\theta)$ whose indices belong to $\mathcal{J}(\theta)$.
$\mathbb{G}_{\mathcal{J}(\theta)}$:	The $J(\theta) \times 1$ vector that stacks components of $\mathbb{G}$ whose indices belong to $\mathcal{J}(\theta)$.
$W_{\mathcal{J}(\theta)}$:	$J(\theta) \times J(\theta)$ matrix that collects (i,j) elements of $W(\theta)$ for $i,j \in \mathcal{J}(\theta)$.

Appendix B: Local Process Regularity

In this Appendix, we first give additional regularity conditions on the local criterion function used to prove Theorem 3.1.

Assumption B.1 (Local Process Regularity)**.** *(i) For any $\epsilon > 0$ and $(u,p) \in \mathbb{R} \times \mathbb{S}^{d-1}$,*

$$P\Big(|\inf_{R_{n,u,p}} \ell_n(\theta,\lambda) - \inf_{R_{u,p}} \ell_n(\theta,\lambda)| \geq \epsilon\Big) \leq \epsilon$$

for n sufficiently large, where $R_{u,p} := H(p,\Theta_I) \times K_{u,p}$. (ii) ℓ_n converges to a convex lower semicontinuous function ℓ_∞ in the following sense; for any $(u_j,p_j), j = 1,\cdots,m$

$$\liminf_{n\to\infty} P(\inf_{R_{u_1,p_1}} \ell_n(\theta,\lambda) > t,\cdots,\inf_{R_{u_m,p_m}} \ell_n(\theta,\lambda) > t) \geq P(\inf_{R_{u_1,p_1}} \ell_\infty(\theta,\lambda) > t,\cdots,\inf_{R_{u_m,p_m}} \ell_\infty(\theta,\lambda) > t) \tag{B.1}$$

$$\limsup_{n\to\infty} P(\inf_{R^o_{u_1,p_1}} \ell_n(\theta,\lambda) \geq t,\cdots,\inf_{R^o_{u_m,p_m}} \ell_n(\theta,\lambda) \geq t) \leq P(\inf_{R^o_{u_1,p_1}} \ell_\infty(\theta,\lambda) \geq t,\cdots,\inf_{R^o_{u_m,p_m}} \ell_\infty(\theta,\lambda) \geq t), \tag{B.2}$$

where $R^o_{u,p} := H(p,\Theta_I) \times K^o_{u,p}$, and $K^o_{u,p} := \{\lambda : \langle p,\lambda\rangle > u\}$. (iii) For each $(u,p) \in \mathbb{R} \times \mathbb{S}^{d-1}$, $\ell_\infty(\theta,\cdot)$ achieves its minimum on $R_{u,p}$. For each $t \in \mathbb{R}_+$ and $p \in \mathbb{S}^{d-1}$, the set $\Lambda_{\theta,t} \equiv \{\lambda : \ell_\infty(\theta,\lambda) \leq t\}$ satisfies $\sup_{\theta \in H(p,\Theta_I)} s(p,\Lambda_{\theta,t}) < \infty$.

Assumption B.1 (i) requires the sequence of sets $R_{n,u,p}$ to converge to a limit $R_{u,p}$. This is satisfied, for example, if the identified set is in the interior of the parameter space. Assumption B.1 (ii) gives the precise notion of convergence required for ℓ_n, which adapts the concept of weak epiconvergence in Knight (1999) and Molchanov (2005). This assumption is satisfied, for example, if the infimum of ℓ_n and ℓ_∞ over $R_{u,p}$ is approximated by its infimum over some compact subset $\tilde{R}_{u,p} \subset R_{u,p}$ and $\inf_{\tilde{R}_{u,p}} \ell_n$ converges weakly to $\inf_{\tilde{R}_{u,p}} \ell_\infty$. Details on the relationship between Assumption B.1 (ii) and other convergence concepts are discussed in the online addendum. Assumption B.1 (iii) requires ℓ_∞'s minimum on $R_{u,p}$ and $\langle p,\theta\rangle$'s maximum over $\Lambda_{t,\theta}$ to be well-defined.

Appendix C: Proof of Theorem 3.1

In this appendix, we establish Theorem 3.1 in multiple steps, which we outline below.

Step 1: We first establish a duality relation between the support function $s(\cdot,\hat{\Theta}_n)$ and the sample criterion function Q_n (Lemma C.1) ■

Step 2: Using Lemma C.1, we then show that the finite-dimensional limit of the normalized support function $\mathcal{Z}_n(p,t) = a_n^{1/\gamma}(s(p,\hat{\Theta}_n) - s(p,\Theta_I))$ can be related to that of $\inf_{(\theta,\lambda)\in R_{u,p}} \ell_n(\theta,\lambda)$ (Lemma C.2) ■

Step 3: In Lemma C.3, we further show that the finite-dimensional distribution of the limiting localized function $\inf_{(\theta,\lambda)\in R_{u,p}} \ell_\infty(\theta,\lambda)$ can be related to that of the limiting process $\mathcal{Z}(p,t)$ ■

Step 4: Combining Steps 2-3, we then show $\mathcal{Z}_n(\cdot,t)$ converges weakly in finite dimension to $\mathcal{Z}(\cdot,t)$. We further strengthen this convergence to weak convergence in $\mathcal{C}(\mathbb{S}^{d-1})$ using Lemma C.4 ■

Lemma C.1 (Duality 1)**.** *Suppose that Assumption 2.1 holds. Let $n \in \mathbb{N}$ and $t \in \mathbb{R}_+$ be given. Then, for any $u \in \mathbb{R}$ and $p \in \mathbb{S}^{d-1}$*

$$s(p,\hat{\Theta}_n(t)) < u \quad \Leftrightarrow \quad \inf_{\theta \in K_{u,p}\cap\Theta} a_n Q_n(\theta) > t \tag{C.1}$$

$$s(p,\hat{\Theta}_n(t)) \leq u \quad \Rightarrow \quad \inf_{\theta \in K^o_{u,p}\cap\Theta} a_n Q_n(\theta) \geq t \tag{C.2}$$

with probability 1, where $K_{u,p}$ is the half space $K_{u,p} := \{\theta \in \mathbb{R}^d : \langle p,\theta\rangle \geq u\}$.

Proof of Lemma C.1. The equivalence (C.1) holds trivially when $K_{u,p} \cap \Theta = \emptyset$ because then the half space $\{\theta \in \mathbb{R}^d : \langle p,\theta\rangle < u\}$ contains Θ and hence also contains $\hat{\Theta}_n(t)$, which in turn implies that $s(p,\hat{\Theta}_n(t)) < u$ must be true. Further, the statement $\inf_{\theta\in\emptyset} a_n Q_n(\theta) = \infty > t$ is always true. Hence (C.1) holds. Below, we assume $K_{u,p} \cap \Theta \neq \emptyset$.

It then follows that

$$
\begin{aligned}
s(p, \hat{\Theta}_n(t)) < u &\Leftrightarrow \langle p, \theta \rangle < u, \quad \forall \theta \in \hat{\Theta}_n(t) \\
&\Leftrightarrow \hat{\Theta}_n(t) \subseteq \Theta \setminus K_{u,p} \\
&\Leftrightarrow K_{u,p} \cap \Theta \subseteq \Theta \setminus \hat{\Theta}_n(t) \\
&\Leftrightarrow a_n Q_n(\theta) > t, \quad \forall \theta \in K_{u,p} \cap \Theta \\
&\Leftrightarrow \inf_{\theta \in K_{u,p} \cap \Theta} a_n Q_n(\theta) > t,
\end{aligned}
\tag{C.3}
$$

where the first equivalence holds because sufficiency is immediate from $s(p, \hat{\Theta}_n(t)) = \sup_{\theta \in \hat{\Theta}_n(t)} \langle p, \theta \rangle$, and the necessity follows from the maximum being achieved on $\hat{\Theta}_n(t)$ by compactness of $\hat{\Theta}_n(t)$ ensured by Assumption 2.1. Similarly, the last equivalence follows from Q_n being lower semicontinuous with probability 1 and $K_{u,p} \cap \Theta$ being a nonempty compact set.

Similarly, (C.2) holds trivially when $K^o_{u,p} \cap \Theta = \emptyset$. Assuming $K^o_{u,p} \cap \Theta \neq \emptyset$ and arguing as in (C.3), it follows that

$$
s(p, \hat{\Theta}_n(t)) \leq u \Leftrightarrow a_n Q_n(\theta) > t, \quad \forall \theta \in K^o_{u,p} \cap \Theta. \tag{C.4}
$$

Since t is a lower bound for the set $\{a_n Q_n(\theta) : \theta \in K^o_{u,p} \cap \Theta\}$, the right hand side of (C.4) implies $\inf_{\theta \in K^o_{u,p} \cap \Theta} a_n Q_n(\theta) \geq t$. Hence, (C.2) holds. This establishes the claim of the Lemma. ■

Lemma C.2 (Duality 2)**.** *Suppose that Assumptions 2.1 and B.1 (i) hold. Let $t \in \mathbb{R}_+$ be given. Then, for any finite m-tuple $\{(u_j, p_j) \in \mathbb{R} \times \mathbb{S}^{d-1}\}_{j=1}^m$,*

$$
\liminf_{n \to \infty} P(\mathcal{Z}_n(p_1, t) < u_1, ..., \mathcal{Z}_n(p_m, t) < u_m)
\geq \liminf_{n \to \infty} P\Big(\inf_{(\theta,\lambda) \in R_{u_1,p_1}} \ell_n(\theta, \lambda) > t, ..., \inf_{(\theta,\lambda) \in R_{u_m,p_m}} \ell_n(\theta, \lambda) > t \Big) \tag{C.5}
$$

$$
\limsup_{n \to \infty} P(\mathcal{Z}_n(p_1, t) \leq u_1, ..., \mathcal{Z}_n(p_m, t) \leq u_m)
\leq \limsup_{n \to \infty} P\Big(\inf_{(\theta,\lambda) \in R^o_{u_1,p_1}} \ell_n(\theta, \lambda) \geq t, ..., \inf_{(\theta,\lambda) \in R^o_{u_m,p_m}} \ell_n(\theta, \lambda) \geq t \Big). \tag{C.6}
$$

Proof of Lemma C.2. We first note that the following equivalence relations hold:

$$
\begin{aligned}
\mathcal{Z}_n(p, t) < u &\Leftrightarrow s(p, \hat{\Theta}_n(t)) < s(p, \Theta_I) + u/a_n^{1/\gamma} \\
&\Leftrightarrow \inf_{\theta' \in K_{s(p,\Theta_I)+u/a_n^{1/\gamma},p} \cap \Theta} a_n Q_n(\theta') > t \Leftrightarrow \inf_{\theta \in H(p,\Theta_I), \lambda \in K_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta)} a_n Q_n(\theta + \lambda/a_n^{1/\gamma}) > t \ ,
\end{aligned}
\tag{C.7}
$$

where the second equivalence follows from Lemma C.1, and the third equivalence follows from the following equality

$$
K_{s(p,\Theta_I)+u/a_n^{1/\gamma},p} \cap \Theta = \{\theta + \lambda/a_n^{1/\gamma} : \theta \in H(p, \Theta_I), \lambda \in K_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta)\}. \tag{C.8}
$$

We show (C.8) below. First, suppose $\theta \in H(p, \Theta_I)$ and $\lambda \in K_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta)$. Then,

$$
\langle p, \theta + \lambda/a_n^{1/\gamma} \rangle = \langle p, \theta \rangle + a_n^{-1/\gamma} \langle p, \lambda \rangle \geq s(p, \Theta_I) + u/a_n^{1/\gamma},
$$

where the last inequality follows from $\theta \in H(p, \Theta_I)$ and $\lambda \in K_{u,p}$. Further, $\theta + \lambda/a_n^{1/\gamma} \in \Theta$ because $\lambda \in a_n^{1/\gamma}(\Theta - \theta)$. Hence, $\theta + \lambda/a_n^{1/\gamma} \in K_{s(p,\Theta_I)+u/a_n^{1/\gamma},p} \cap \Theta$. This establishes that the set on the right hand side of (C.8) is a subset of $K_{s(p,\Theta_I)+u/a_n^{1/\gamma},p} \cap \Theta$. For the reverse inclusion, let $\theta' \in K_{s(p,\Theta_I)+u/a_n^{1/\gamma},p} \cap \Theta$ and pick any $\theta \in H(p, \Theta_I)$. We then let $\lambda := a_n^{1/\gamma}(\theta' - \theta)$. By construction, we have $\lambda \in a_n^{1/\gamma}(\Theta - \theta)$. Further,

$$
\langle p, \lambda \rangle = \langle p, a_n^{1/\gamma}(\theta' - \theta) \rangle = a_n^{1/\gamma}(\langle p, \theta' \rangle - \langle p, \theta \rangle) \geq a_n^{1/\gamma}(s(p, \Theta_I) + u/a_n^{1/\gamma} - s(p, \Theta_I)) = u,
$$

where the inequality follows from $\theta' \in K_{s(p,\Theta_I)+u/a_n^{1/\gamma}} \cap \Theta$. Therefore, the reverse inclusion holds. This in turn establishes (C.8).

Now, by (C.7) and the definition of ℓ_n and $R_{n,u,p}$, we obtain $\mathcal{Z}_n(p,t) < u \Leftrightarrow \inf_{R_{n,u,p}} \ell_n(\theta,\lambda) > t$. Since this holds for any finite m-tuple $\{(u_j,p_j)\}_{j=1}^m$, it follows that

$$P(\mathcal{Z}_n(p_1,t) < u_1,\cdots,\mathcal{Z}_n(p_m,t) < u_m) = P\Big(\inf_{R_{n,u_1,p_1}} \ell_n(\theta,\lambda) > t,\cdots,\inf_{R_{n,u_m,p_m}} \ell_n(\theta,\lambda) > t\Big). \tag{C.9}$$

Note that, for any $\epsilon > 0$, we have

$$\begin{aligned} &P\Big(\inf_{R_{u_1,p_1}} \ell_n(\theta,\lambda) > t+\epsilon,\cdots,\inf_{R_{u_m,p_m}} \ell_n(\theta,\lambda) > t+\epsilon\Big) \\ &\quad\le P\Big(\max_{1\le j\le m}\Big|\inf_{R_{u_j,p_j}} \ell_n(\theta,\lambda) - \inf_{R_{n,u_j,p_j}} \ell_n(\theta,\lambda)\Big| \ge \epsilon\Big) + P\Big(\inf_{R_{n,u_1,p_1}} \ell_n(\theta,\lambda) > t,\cdots,\inf_{R_{n,u_m,p_m}} \ell_n(\theta,\lambda) > t\Big), \end{aligned} \tag{C.10}$$

where we used the fact that, for any random vectors $Y_n, X_n : \Omega \to \mathbb{R}^m$, an open set $G \subset \mathbb{R}^m$, and its ϵ-contraction $G^{-\epsilon} := \{x \in G : \rho(x, G^c) \ge \epsilon\}$, we have $P(Y_n \in G^{-\epsilon}) \le P(\rho(X_n,Y_n) \ge \epsilon) + P(X_n \in G)$. Specifically, we used the metric $\rho(X_n,Y_n) = \max_{1\le j\le m} |X_{j,n} - Y_{j,n}|$ and the open set $G = (t,\infty)^m$. Assumption B.1 (i) ensures that the first term on the right hand side of (C.10) becomes arbitrarily small as n gets large. Therefore,

$$\begin{aligned} &\liminf_{n\to\infty} P\Big(\inf_{R_{u_1,p_1}} \ell_n(\theta,\lambda) > t+\epsilon,\cdots,\inf_{R_{u_m,p_m}} \ell_n(\theta,\lambda) > t+\epsilon\Big) \\ &\qquad\qquad \le \liminf_{n\to\infty} P\Big(\inf_{R_{n,u_1,p_1}} \ell_n(\theta,\lambda) > t,\cdots,\inf_{R_{n,u_m,p_m}} \ell_n(\theta,\lambda) > t\Big). \end{aligned} \tag{C.11}$$

Since ϵ is arbitrary, (C.5) then follows from (C.9) and (C.11).

Similarly, by Lemma C.1 and an argument as in (C.7), it follows that $\mathcal{Z}_n(p,t) \le u \Rightarrow \inf_{R^o_{n,u,p}} \ell_n(\theta,\lambda) \ge t$. Since this holds for any finite m-tuple $\{(u_j,p_j)\}_{j=1}^m$, we have

$$P(\mathcal{Z}_n(p_1,t) \le u_1,\cdots,\mathcal{Z}_n(p_m,t) \le u_m) \le P\Big(\inf_{R^o_{n,u_1,p_1}} \ell_n(\theta,\lambda) \ge t,\cdots,\inf_{R^o_{n,u_m,p_m}} \ell_n(\theta,\lambda) \ge t\Big). \tag{C.12}$$

Note that, for any $\epsilon > 0$, we have

$$\begin{aligned} &P\Big(\inf_{R^o_{n,u_1,p_1}} \ell_n(\theta,\lambda) \ge t,\cdots,\inf_{R^o_{n,u_m,p_m}} \ell_n(\theta,\lambda) \ge t\Big) \\ &\quad\le P\Big(\max_{1\le j\le m}\Big|\inf_{R^o_{u_j,p_j}} \ell_n(\theta,\lambda) - \inf_{R^o_{n,u_j,p_j}} \ell_n(\theta,\lambda)\Big| \ge \epsilon\Big) + P\Big(\inf_{R^o_{u_1,p_1}} \ell_n(\theta,\lambda) \ge t-\epsilon,\cdots,\inf_{R^o_{u_m,p_m}} \ell_n(\theta,\lambda) \ge t-\epsilon\Big), \end{aligned} \tag{C.13}$$

where we used the fact that, for any random vectors $Y_n, X_n : \Omega \to \mathbb{R}^m$, a closed set $F \subset \mathbb{R}^m$, and its ϵ-expansion $F^\epsilon := \{x \in F : \rho(x,F) \le \epsilon\}$, we have $P(Y_n \in F) \le P(\rho(X_n,Y_n) \ge \epsilon) + P(X_n \in F^\epsilon)$. By Assumption B.1 (i),

$$\begin{aligned} &\limsup_{n\to\infty} P\Big(\inf_{R^o_{n,u_1,p_1}} \ell_n(\theta,\lambda) \ge t,\cdots,\inf_{R^o_{n,u_m,p_m}} \ell_n(\theta,\lambda) \ge t\Big) \\ &\qquad\qquad \le \limsup_{n\to\infty} P\Big(\inf_{R^o_{u_1,p_1}} \ell_n(\theta,\lambda) \ge t-\epsilon,\cdots,\inf_{R^o_{u_m,p_m}} \ell_n(\theta,\lambda) \ge t-\epsilon\Big). \end{aligned} \tag{C.14}$$

Since ϵ is arbitrary, (C.6) then follows from (C.12) and (C.14). This completes the proof. ■

Lemma C.3. *Suppose Assumption B.1 holds. Let $t \in \mathbb{R}_+$ be given. Then, for any finite m-tuple $\{(u_j,p_j) \in \mathbb{R}\times\mathbb{S}^{d-1}\}_{j=1}^m$,*

$$P\Big(\mathcal{Z}(p_1,t) < u_1,\cdots,\mathcal{Z}(p_m,t) < u_m\Big) \le P\Big(\inf_{(\theta,\lambda)\in R_{u_1,p_1}} \ell_\infty(\theta,\lambda) > t,\cdots,\inf_{(\theta,\lambda)\in R_{u_m,p_m}} \ell_\infty(\theta,\lambda) > t\Big) \tag{C.15}$$

$$P\Big(\mathcal{Z}(p_1,t) \le u_1,\cdots,\mathcal{Z}(p_m,t) \le u_m\Big) \ge P\Big(\inf_{(\theta,\lambda)\in R^o_{u_1,p_1}} \ell_\infty(\theta,\lambda) \ge t,\cdots,\inf_{(\theta,\lambda)\in R^o_{u_m,p_m}} \ell_\infty(\theta,\lambda) \ge t\Big). \tag{C.16}$$

Proof of Lemma C.3. Let $p \in \mathbb{S}^{d-1}$. We first note that $\sup_{\theta\in H(p,\Theta_I)} s(p,\Lambda_{\theta,t}) < \infty$ by Assumption B.1 (iii). Arguing as in (C.3), we then obtain

$$
\begin{aligned}
\sup_{\theta\in H(p,\Theta_I)} s(p,\Lambda_{\theta,t}) < u \Rightarrow \langle p,\lambda\rangle < u,\ \forall\theta\in H(p,\Theta_I) \text{ and } \lambda\in\Lambda_{\theta,t},\\
&\Leftrightarrow \Lambda_{\theta,t}\subseteq\mathbb{R}^d\setminus K_{u,p},\ \forall\theta\in H(p,\Theta_I)\\
&\Leftrightarrow K_{u,p}\subseteq\mathbb{R}^d\setminus\Lambda_{\theta,t},\ \forall\theta\in H(p,\Theta_I)\\
&\Leftrightarrow \ell_\infty(\theta,\lambda) > t,\ \forall\theta\in H(p,\Theta_I) \text{ and } \lambda\in K_{u,p}\\
&\Leftrightarrow \inf_{(\theta,\lambda)\in R_{u,p}} \ell_\infty(\theta,\lambda) > t,
\end{aligned}
\tag{C.17}
$$

where the last equivalence follows from Assumption B.1 (iii). Since this holds for any finite m-tuple $\{(u_j,p_j), j = 1,\cdots,m\}$, (C.15) holds.

Let $\Lambda^o_{\theta,t} := \{\lambda\in\mathbb{R}^d : \ell_\infty(\theta,\lambda) < t\}$. We then have

$$
\begin{aligned}
\inf_{(\theta,\lambda)\in R^o_{u_1,p_1}} \ell_\infty(\theta,\lambda)\ge t \Rightarrow \ell_\infty(\theta,\lambda)\ge t,\ \forall\theta\in H(p,\Theta_I) \text{ and } \lambda\in K^o_{u,p}\\
&\Leftrightarrow \Lambda^o_{\theta,t}\subseteq\mathbb{R}^d\setminus K^o_{u,p},\ \forall\theta\in H(p,\Theta_I)\\
&\Rightarrow s(p,\Lambda_{\theta,t})\le u,\ \forall\theta\in H(p,\Theta_I)\\
&\Rightarrow \sup_{\theta\in H(p,\Theta_I)} s(p,\Lambda_{\theta,t})\le u,
\end{aligned}
\tag{C.18}
$$

where the second equivalence follows from the definition of $\Lambda^o_{\theta,t}$, the second implication follows because $\mathbb{R}^d\setminus K^o_{u,p} = \{\lambda : \langle p,\lambda\rangle\le u\}$ is closed hence contains $\Lambda_{\theta,t} = \mathrm{cl}(\Lambda^o_{\theta,t})$, implying $\Lambda_{\theta,t}$'s support function being weakly dominated by u. Since this holds for any finite m-tuple $\{(u_j,p_j), j = 1,\cdots,m\}$, (C.16) holds. ∎

Lemma C.4. *Let $\mathbb{E}$ be a compact set in a metric space. Let $h : [0,\infty)\to[0,\infty)$ be a function such that $h(0) = 0$ and h is continuous at 0. There is B_n such that $B_n = O_p(1)$. If for all $x,y\in\mathbb{E}$, $|\xi_n(x)-\xi_n(y)|\le B_n h(\|x-y\|)$, then $\{\xi_n\}$ is stochastically equicontinuous.*

Proof of Lemma C.4. The result immediately follows from Assumption 3A and Corollary 2.2 in Newey (1991). ∎

Proof of Theorem 3.1. By Assumption B.1 (ii), Lemmas C.2 and C.3, it follows that, for any $\{(u_j,p_j)\}_{j=1}^m$,

$$
\begin{aligned}
&\liminf_{n\to\infty} P(\mathcal{Z}_n(p_1,t) < u_1, ..., \mathcal{Z}_n(p_m,t) < u_m)\\
&\qquad\ge \liminf_{n\to\infty} P\Big(\inf_{(\theta,\lambda)\in R_{u_1,p_1}} \ell_n(\theta,\lambda) > t, ..., \inf_{(\theta,\lambda)\in R_{u_m,p_m}} \ell_n(\theta,\lambda) > t\Big)\\
&\qquad\ge P\Big(\inf_{(\theta,\lambda)\in R_{u_1,p_1}} \ell_\infty(\theta,\lambda) > t,\cdots, \inf_{(\theta,\lambda)\in R_{u_m,p_m}} \ell_\infty(\theta,\lambda) > t\Big)\\
&\qquad\ge P\Big(\mathcal{Z}(p_1,t) < u_1,\cdots,\mathcal{Z}(p_m,t) < u_m\Big).
\end{aligned}
\tag{C.19}
$$

Similarly, by Assumption B.1 (ii), Lemmas C.2 and C.3, it follows that

$$
\begin{aligned}
&\limsup_{n\to\infty} P(\mathcal{Z}_n(p_1,t)\le u_1, ...,\mathcal{Z}_n(p_m,t)\le u_m)\\
&\qquad\le \limsup_{n\to\infty} P\Big(\inf_{(\theta,\lambda)\in R^o_{u_1,p_1}} \ell_n(\theta,\lambda)\ge t, ..., \inf_{(\theta,\lambda)\in R^o_{u_m,p_m}} \ell_n(\theta,\lambda)\ge t\Big)\\
&\qquad\le P\Big(\inf_{(\theta,\lambda)\in R^o_{u_1,p_1}} \ell_\infty(\theta,\lambda)\ge t,\cdots, \inf_{(\theta,\lambda)\in R^o_{u_m,p_m}} \ell_\infty(\theta,\lambda)\ge t\Big)\\
&\qquad\le P\Big(\mathcal{Z}(p_1,t)\le u_1,\cdots,\mathcal{Z}(p_m,t)\le u_m\Big).
\end{aligned}
\tag{C.20}
$$

By (C.19) and (C.20), it follows that for any continuity point $(u_1, \cdots, u_m)$ of $(\mathcal{Z}(p_1,t), \cdots, \mathcal{Z}(p_m,t))'$,

$$\begin{aligned}\limsup_{n\to\infty} P(\mathcal{Z}_n(p_1,t) \le u_1, ..., \mathcal{Z}_n(p_m,t) \le u_m) &\le P(\mathcal{Z}(p_1,t) \le u_1, \cdots, \mathcal{Z}(p_m,t) \le u_m) \\ &= P(\mathcal{Z}(p_1,t) < u_1, \cdots, \mathcal{Z}(p_m,t) < u_m) \\ &\le \liminf_{n\to\infty} P(\mathcal{Z}_n(p_1,t) < u_1, ..., \mathcal{Z}_n(p_m,t) < u_m) \\ &\le \liminf_{n\to\infty} P(\mathcal{Z}_n(p_1,t) \le u_1, ..., \mathcal{Z}_n(p_m,t) \le u_m). \end{aligned} \tag{C.21}$$

Since $\liminf_{n\to\infty} P(\mathcal{Z}_n(p_1,t) \le u_1, ..., \mathcal{Z}_n(p_m,t) \le u_m) \le \limsup_{n\to\infty} P(\mathcal{Z}_n(p_1,t) \le u_1, ..., \mathcal{Z}_n(p_m,t) \le u_m)$ always holds, (C.21) ensures that $\lim_{n\to\infty} P(\mathcal{Z}_n(p_1,t) \le u_1, ..., \mathcal{Z}_n(p_m,t) \le u_m) = P(\mathcal{Z}(p_1,t) \le u_1, ..., \mathcal{Z}(p_m,t) \le u_m)$ at any continuity point $(u_1, \cdots, u_m)$ of $(\mathcal{Z}(p_1,t), \cdots, \mathcal{Z}(p_m,t))'$. Therefore $\mathcal{Z}_n(\cdot,t)$ converges weakly in finite dimension to $\mathcal{Z}(\cdot,t)$. This establishes (i).

For (ii), it then suffices to show the tightness of $\mathcal{Z}_n(\cdot,t)$. We now show the required conditions for Lemma C.4 using an expansion of the support function. In the following, we extend $s(\cdot,\Theta_I)$ and $s(\cdot,\hat{\Theta}_n)$ from $\mathbb{S}^{d-1}$ to $\mathbb{R}^d$. Under our assumptions, Θ_I is a compact convex set, and $\hat{\Theta}_n(t)$ is a compact convex set almost surely. This ensures that $p \mapsto s(p,\Theta_I)$ is convex, and $p \mapsto s(p,\hat{\Theta}_n(t))$ is convex $a.s.$ Now, take an open convex set O such that $\mathbb{S}^{d-1} \subset O$. Let $p, q \in \mathbb{S}^{d-1}$. Then, by Theorem 10.48 in Rockafellar and Wets (2005), for some $\bar{p}_n$ and $\bar{p}$ on the line segment that connects p and q, there exist $\hat{v}_n \in \partial s(\bar{p}_n, \hat{\Theta}_n(t))$ and $w \in \partial s(\bar{p}, \Theta_I)$ such that

$$s(p,\hat{\Theta}_n(t)) - s(q,\hat{\Theta}_n(t)) = \langle \hat{v}_n, p-q\rangle \tag{C.22}$$

$$s(p,\Theta_I) - s(q,\Theta_I) = \langle w, p-q\rangle \tag{C.23}$$

Subtracting (C.23) from (C.22) and multiplying both sides by $a_n^{1/\gamma}$ yields

$$\mathcal{Z}_n(p,t) - \mathcal{Z}_n(q,t) = a_n^{1/\gamma}\langle \hat{v}_n - w, p-q\rangle. \tag{C.24}$$

Note that, under Assumptions 2.2-2.3, CHT's Theorem 3.2 and Theorem 1.1.12 in Li, Ogura, and Kreinovich (2002) imply $\mathcal{Z}_n(p,t) = O_p(1)$ for any $p \in \mathbb{S}^{d-1}$. Therefore $a_n^{1/\gamma}\langle \hat{v}_n - w, p-q\rangle = \mathcal{Z}_n(p,t) - \mathcal{Z}_n(q,t) = O_p(1)$ for any $p, q \in \mathbb{S}^{d-1}$. Since this holds for any p and q, each component of $a_n^{1/\gamma}(\hat{v}_n - w)$ must be $O_p(1)$. Therefore, $a_n^{1/\gamma}\|\hat{v}_n - w\| = O_p(1)$.

Applying the Cauchy-Schwarz inequality to (C.24), we obtain

$$|\mathcal{Z}_n(p,t) - \mathcal{Z}_n(q,t))| \le a_n^{1/\gamma}\|\hat{v}_n - w\|\|p-q\|.$$

Now, we apply Lemma C.4 with $B_n = a_n^{1/\gamma}\|\hat{v}_n - w\|$ and $h(x) = x$. This ensures that $\{\mathcal{Z}_n(\cdot,t), n \ge 1\}$ is stochastically equicontinuous. Thus, $\{\mathcal{Z}_n(\cdot,t), n \ge 1\}$ is tight. Note that a tight sequence that is weakly converging in finite dimension weakly converges in the uniform metric (van der Vaart and Wellner, 2000). Thus, we obtain $\mathcal{Z}_n(\cdot,t) \overset{u.d.}{\to} \mathcal{Z}(\cdot,t)$. The weak convergence of the Hausdorff distances then follow from (2.7) and the continuous mapping theorem. ■

Appendix D: Proof of Theorems 3.2, 3.3 3.4 and Corollaries 3.1-3.2

In this section, we give the proof of Theorems 3.2, 3.3 and 3.4, Corollaries 3.1, 3.2 and auxiliary lemmas. Theorem 3.4 is proved using a functional δ-method for directionally differentiable functionals. For this, we need a suitable differentiability concept of the map $s(\cdot,\Theta_I) \mapsto \sup_{p\in\mathbb{S}^{d-1}}\langle p,\theta_0\rangle - s(p,\Theta_I)$. The following definition is based on Shapiro (1991).

Definition D.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be normed vector spaces. A map $g : \mathcal{X} \to \mathcal{Y}$ is said to be* Hadamard directionally differentiable *at μ if for every sequence $\{t_n\}$ of positive numbers converging to 0 and any sequence $\{x_n\}$ converging* *to x, the limit*

$$\dot{g}(x) = \lim_{n\to\infty} \frac{g(\mu + t_n x_n) - g(\mu)}{t_n} \tag{D.1}$$

exists. If $\dot{g}$ is linear in x, then g is said to be Hadamard differentiable *at μ.*

Proof of Theorem 3.2. For each $t \in \mathbb{R}_+$ and $p \in \mathbb{S}^{d-1}$, let $\mathcal{Z}^*_{n,b,k}(t) := a_b^{1/\gamma}[s(p, \hat{\Theta}_{n,b,k}(t)) - s(p, \Theta_I)]$. For each $x \in \mathbb{R}$ and $t \in \mathbb{R}_+$, let

$$U_{n,b}(x,t) := N_{n,b}^{-1} \sum_{k=1}^{N_{n,b}} 1\Big\{ \sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}^*_{n,b,k}(p,t)) \le x \Big\}. \tag{D.2}$$

Let $\epsilon, \delta > 0$, and let K be the Lipschitz constant of Υ. Suppose that $\sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) \le x$, $a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \le \epsilon/2K$, $d_H(\hat{\Psi}_n, \Psi_0) \le \delta$, and $\sup_{\|p-p'\|\le\delta} |\mathcal{Z}_{n,b,k}(p,t) - \mathcal{Z}_{n,b,k}(p,t)| \le \epsilon/2K$. Then, it follows that

$$\begin{aligned}
&|\sup_{p\in\hat{\Psi}_n} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}^*_{n,b,k}(p,t))| \\
&\quad \le |\sup_{p\in\hat{\Psi}_n} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}_{n,b,k}(p,t))| + |\sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}^*_{n,b,k}(p,t))| \\
&\quad \le |\sup_{p\in\hat{\Psi}_n} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}_{n,b,k}(p,t))| + \sup_{p\in\mathbb{S}^{d-1}} |\Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \Upsilon(\mathcal{Z}^*_{n,b,k}(p,t))|.
\end{aligned} \tag{D.3}$$

Let $\hat{p}_n \in \arg\max_{p\in\hat{\Psi}_n} \Upsilon(\mathcal{Z}_{n,b,k}(p,t))$, which is well defined by the compactness of $\hat{\Psi}_n$ and the continuity of the map $p \mapsto \Upsilon(\mathcal{Z}_{n,b,k}(p,t))$. Let $\Pi_{\Psi_0}\hat{p}_n$ be the projection of $\hat{p}_n$ on Ψ_0 and note that $\|\hat{p}_n - \Pi_{\Psi_0}\hat{p}_n\| \le d_H(\hat{\Psi}_n, \Psi_0) \le \delta$. We then obtain,

$$\begin{aligned}
&\sup_{p\in\hat{\Psi}_n} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) \\
&\quad \le \Upsilon(\mathcal{Z}_{n,b,k}(\hat{p}_n,t)) - \Upsilon(\mathcal{Z}_{n,b,k}(\Pi_{\Psi_0}\hat{p}_n,t)) \le \sup_{\|p-p'\|\le\delta} |\Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \Upsilon(\mathcal{Z}_{n,b,k}(p',t))|.
\end{aligned} \tag{D.4}$$

A similar argument gives

$$\sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \sup_{p\in\hat{\Psi}_n} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) \le \sup_{\|p-p'\|\le\delta} |\Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \Upsilon(\mathcal{Z}_{n,b,k}(p',t))|. \tag{D.5}$$

(D.4) and (D.5) then imply

$$\begin{aligned}
&|\sup_{p\in\hat{\Psi}_n} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \sup_{p\in\Psi_0} \Upsilon(Z_{n,b,k}(p,t))| \\
&\quad \le \sup_{\|p-p'\|\le\delta} |\Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \Upsilon(\mathcal{Z}_{n,b,k}(p',t))| \le K \sup_{\|p-p'\|\le\delta} |\mathcal{Z}_{n,b,k}(p,t) - \mathcal{Z}_{n,b,k}(p',t)| \le \frac{\epsilon}{2},
\end{aligned} \tag{D.6}$$

by the Lipschitz continuity of Υ and the hypothesis. Furthermore, the Lipschitz continuity of Υ and the hypothesis that $a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \le \epsilon/2K$ ensure

$$\sup_{p\in\mathbb{S}^{d-1}} |\Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \Upsilon(\mathcal{Z}^*_{n,b,k}(p,t))| \le K \sup_{p\in\mathbb{S}^{d-1}} a_b^{1/b} |s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I)| \le \frac{\epsilon}{2}. \tag{D.7}$$

Combining Eqs. (D.3)-(D.7) yields $|\sup_{p\in\hat{\Psi}_n} \Upsilon(\mathcal{Z}_{n,b,k}(p,t)) - \sup_{p\in\Psi_0} \Upsilon(\mathcal{Z}^*_{n,b,k}(p,t))| \le \epsilon$. Therefore, we have $\sup_{p\in\Psi_0} \Upsilon(Z^*_{n,b,k}(p,t)) \le x + \epsilon$. Now define the following event:

$$E_{n,b}(t,\epsilon,\delta) := \Big\{\omega \in \Omega : a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \le \epsilon/2K,\ d_H(\hat{\Psi}_n, \Psi_0) \le \delta,\ \sup_{\|p-p'\|\le\delta} |\mathcal{Z}_{n,b,k}(p,t) - \mathcal{Z}_{n,b,k}(p,t)| \le \epsilon/2K \Big\}. \tag{D.8}$$

The arguments above ensure that the following inequality holds:

$$\hat{F}_{n,b}(x,t) 1_{E_{n,b}(t,\epsilon,\delta)} \le U_{n,b}(x+\epsilon, t). \tag{D.9}$$

Now, suppose on the other hand that $\sup_{p\in\Psi_0}\Upsilon(\mathcal{Z}^*_{n,b,k}(p,t))\le x-\epsilon$ and that $\omega\in E_{n,b}(t,\epsilon,\delta)$. Then, using the same argument as above, it is straightforward to show that

$$|\sup_{p\in\hat{\Psi}_n}\Upsilon(\mathcal{Z}_{n,b,k}(p,t))-\sup_{p\in\Psi_0}\Upsilon(\mathcal{Z}^*_{n,b,k}(p,t))|\le\epsilon. \tag{D.10}$$

Hence, we have $\sup_{p\in\hat{\Psi}_n}\Upsilon(\mathcal{Z}_{n,b,k}(p,t))\le x$. Therefore, we obtain the following inequality:

$$U_{n,b}(x-\epsilon,t)1_{E_{n,b}(t,\epsilon,\delta)}\le\hat{F}_{n,b}(x,t)1_{E_{n,b}(t,\epsilon,\delta)}. \tag{D.11}$$

By CHT's Theorem 3.1 (1), the assumption that $d_H(\hat{\Psi}_n,\Psi_0)=o_p(1)$, and the stochastic equicontinuity of $\{\mathcal{Z}_{n,b,k}(\cdot,t)\}$, (D.9) and (D.11) hold for any $\epsilon,\delta>0$ and $P(E_{n,b}(t,\epsilon,\delta))\to1$ as $n\to\infty$ and $b\to\infty$. Hence, for any $\epsilon>0$, we have

$$U_{n,b}(x-\epsilon,t)\le\hat{F}_{n,b}(x,t)\le U_{n,b}(x+\epsilon,t), \tag{D.12}$$

with probability tending to 1. Now it is straightforward to show $U_{n,b}(x-\epsilon,t)=F(x,t)+o_p(1)$ for each continuity point x of $F(\cdot,t)$ by an argument similar to the proof of Theorem 2.2.1 (i) in Politis, Romano, and Wolf (1999). Therefore,

$$F(x-\epsilon,t)-\epsilon\le\hat{F}_{n,b}(x,t)\le F(x+\epsilon,t)+\epsilon,$$

with probability tending to 1 for any $\epsilon>0$. Now, let $\epsilon\downarrow0$ so that $x\pm\epsilon$ are continuity points of $F^{\rightarrow}(\cdot,P)$. Then, the conclusion follows. ■

Proof of Corollary 3.1. (i) Let $\hat{F}^{\rightarrow}_n(\cdot,t)$ be the empirical cdf of $T^{\rightarrow}_n(t)$. Similarly, for each t, let $F^{\rightarrow}(\cdot,t)$ be the cdf of $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+$. Let $\Upsilon(x)=\{-x\}_+$ and $\Psi_0=\Psi_n=\mathbb{S}^{d-1}$. By Theorem 3.2, it follows that $\hat{F}^{\rightarrow}_n(x,t)-F^{\rightarrow}(x,t)=o_p(1)$ at every continuity point of $F^{\rightarrow}(\cdot,t)$.

Note that the map $x\mapsto F^{\rightarrow}(x,t)$ is right continuous with a left limit. Hence, for any $\epsilon>0$, there is $\eta_\epsilon>0$ such that

$$|F^{\rightarrow}(x,t)-F^{\rightarrow}(c^{\rightarrow}_{1-\alpha}(t),t)|<\epsilon,\ \forall x\in(c^{\rightarrow}_{1-\alpha}(t),\ c^{\rightarrow}_{1-\alpha}(t)+\eta_\epsilon) \tag{D.13}$$

$$|F^{\rightarrow}(x,t)-L|<\epsilon,\ \forall x\in(c^{\rightarrow}_{1-\alpha}(t)-\eta_\epsilon,\ c^{\rightarrow}_{1-\alpha}(t))\ , \tag{D.14}$$

where $L=\lim_{c\uparrow c^{\rightarrow}_{1-\alpha}(t)}F^{\rightarrow}(c,t)$ is the left limit of $F^{\rightarrow}$ at $c^{\rightarrow}_{1-\alpha}(t)$. Hence, there are continuity points c_0,c_1 of $F^{\rightarrow}(\cdot,t)$ such that $c_1<c^{\rightarrow}_{1-\alpha}(t)<c_0$. Below, we take c_0 and c_1 such that $c_0-c_1=\delta$, where δ is the constant specified in the corollary. By $c_1<c^{\rightarrow}_{1-\alpha}(t)$, it must be the case that

$$F^{\rightarrow}(c_1,t)<1-\alpha, \tag{D.15}$$

because otherwise we must have $c_1\ge c^{\rightarrow}_{1-\alpha}(t)$ by $c^{\rightarrow}_{1-\alpha}(t)$ being the infimum of c such that $F^{\rightarrow}(c,t)\ge1-\alpha$. Since c_1 is a continuity point of $F^{\rightarrow}(\cdot,t)$ and $\hat{F}^{\rightarrow}_n(x,t)-F^{\rightarrow}(x,t)=o_p(1)$ at each continuity point of $F^{\rightarrow}(\cdot,t)$, it then follows that

$$\hat{F}^{\rightarrow}_n(c_1,t)<1-\alpha\ ,\quad wp\to1\ . \tag{D.16}$$

This in turn implies

$$c_1<\inf\{c\in\mathbb{R}:\hat{F}^{\rightarrow}_n(c,t)\ge1-\alpha\}=\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t),\quad wp\to1. \tag{D.17}$$

For each n, let A_n be defined by

$$A_n:=\{\omega\in\Omega:c_1<\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\}=\{\omega\in\Omega:c_0<\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)+\delta\}\ , \tag{D.18}$$

where the second equality follows from $c_1=c_0-\delta$. Now note that:

$$P(T^{\rightarrow}_n(t)>\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)+\delta)\le P(\{T^{\rightarrow}_n(t)>\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)+\delta\}\cap A_n)+P(A^c_n)\le P(T^{\rightarrow}_n(t)>c_0)+P(A^c_n). \tag{D.19}$$

Then, we obtain

$$\begin{aligned}\limsup_{n\to\infty} P(T_n^{\to}(t) > \hat{c}_{n,b,1-\alpha}^{\to}(t)+\delta) &\le \limsup_{n\to\infty} P(T_n^{\to}(t) > c_0) + \limsup_{n\to\infty} P(A_n^c)\\ &\le \limsup_{n\to\infty} P(T_n^{\to}(t) \ge c_0) + \limsup_{n\to\infty} P(A_n^c)\\ &\overset{(1)}{\le} P(\sup_{p\in\mathbb{S}^l}\{-\mathcal{Z}(p,t)\}_+ \ge c_0)\\ &\overset{(2)}{=} P(\sup_{p\in\mathbb{S}^l}\{-\mathcal{Z}(p,t)\}_+ > c_0) = 1-F^{\to}(c_0,t) \le 1-F^{\to}(c_{1-\alpha}^{\to}(t),t) \le \alpha. \end{aligned} \quad \text{(D.20)}$$

where (1) follows from $T_n^{\to}(t) \xrightarrow{d} \sup_{p\in\mathbb{S}^l}\{-\mathcal{Z}(p,t)\}_+$ and $P(A_n^c)\to 0$ by (D.17)-(D.18). Equality (2) follows from c_0 being a continuity point of $F^{\to}(\cdot,t)$, and the rest follows from $c_0 > c_{1-\alpha}^{\to}(t)$ and the definition of $c_{1-\alpha}^{\to}(t)$.

Part (ii) follows from the fact that $T_n^{\to}(t) = \sup_{p\in\mathbb{S}^l}\{-\mathcal{Z}_n(p,t) + a_n^{1/\gamma}(s(p,\Theta_0)-s(p,\Theta_I))\}_+ \xrightarrow{p} \infty$ under a fixed alternative and that $\tilde{c}_{n,b,1-\alpha}(t) = O_p(1)$. ■

We use the following two lemmas (Lemmas D.1 and D.2) to show Theorem 3.3.

Lemma D.1. *Suppose the conditions of Theorem 3.1 are satisfied. Then, for any* $0\le t < t' \le t^*_{1-\alpha}$*,*

$$0 = c_{1-\alpha}^{\to}(t^*_{1-\alpha}) \le c_{1-\alpha}^{\to}(t') \le c_{1-\alpha}^{\to}(t) \le c_{1-\alpha}^{\to}(0). \quad \text{(D.21)}$$

Proof of Lemma D.1. First, $c_{1-\alpha}^{\to}(t^*_{1-\alpha}) = 0$ follows from the definition of $t^*_{1-\alpha}$. For the conclusion of the lemma, it suffices to show that $P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \le x\right)$ is non-decreasing in t for each x. As this is a distributional property of the process $\mathcal{Z}(p,t)$, it suffices to show that the statement above holds for the following representation:

$$-\mathcal{Z}(p,t) = -\sup_{\theta\in H(p,\Theta_I)}\ \sup_{\lambda\in\{\lambda:\ell_\infty(\theta,\lambda)\le t\}} \langle p,\lambda\rangle. \quad \text{(D.22)}$$

As $\{\lambda:\ell_\infty(\theta,\lambda)\le t\}\subseteq\{\lambda:\ell_\infty(\theta,\lambda)\le t'\}$ for any $0\le t<t'\le t^*_{1-\alpha}$ and for each $p\in\mathbb{S}^{d-1}$, $-\mathcal{Z}(p,t)$ is non-increasing in t. This implies that $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+$ is non-increasing in t for any ω. Thus, $P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+\le x\right)$ is non-decreasing in $t\in[0,t^*_{1-\alpha}]$ for each x. ■

Lemma D.2. *Suppose the conditions of Theorem 3.3 hold. Then, for any* $\alpha\in(0,1)$ *and* $0\le t<t'\le t^*_{1-\alpha}$*,* $c_{1-\alpha}^{\to}(t) - c_{1-\alpha}^{\to}(t') = \mu(t')-\mu(t)$*.*

Proof of Lemma D.2. First, $c_{1-\alpha}^{\to}(t)$ can be written as

$$\begin{aligned}c_{1-\alpha}^{\to}(t) &= \inf\left\{x : P\Big(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+\le x\Big)\ge 1-\alpha\right\}\\ &= \inf\left\{x : P\Big(\sup_{p\in\mathbb{S}^{d-1}}\{\mu(t')-\mu(t)-\mu(t')-\mathcal{Z}^*(p)\}_+\le x\Big)\ge 1-\alpha\right\}. \end{aligned} \quad \text{(D.23)}$$

Let $\Delta(t,t') := \mu(t')-\mu(t)$. Then, for any $x\ge\Delta(t,t')$, it follows that

$$\begin{aligned}&P\Big(\sup_{p\in\mathbb{S}^{d-1}}\{\mu(t')-\mu(t)-\mu(t')-\mathcal{Z}^*(p)\}_+\le x\Big)\\ &\quad= P\Big(\sup_{p\in\mathbb{S}^{d-1}}\{\Delta(t,t')-\mathcal{Z}(p,t')\}_+\le x\Big) = P\Big(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t')\}_+\le x-\Delta(t,t')\Big). \end{aligned} \quad \text{(D.24)}$$

Substituting Eq. (D.24) into Eq. (D.23) yields

$$c_{1-\alpha}^{\to}(t) = \inf\left\{x : P\Big(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t')\}_+\le x-\Delta(t,t')\Big)\ge 1-\alpha\right\} = c_{1-\alpha}^{\to}(t') + \Delta(t,t'). \quad \text{(D.25)}$$

This establishes the claim of the lemma. ■

Proof of Theorem 3.3. By Theorem in 1.1.12 in Li, Ogura, and Kreinovich (2002),

$$\begin{aligned}
a_n^{1/\gamma} d_H\big(\mathcal{C}_{1n}(t), \hat{\Theta}_n(t^*_{1-\alpha})\big) &= a_n^{1/\gamma} \sup_{p\in\mathbb{S}^{d-1}} |s(p,\hat{\Theta}_n(t)) + \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)/a_n^{1/\gamma} - s(p,\hat{\Theta}_n(t^*_{1-\alpha}))| \\
&= \sup_{p\in\mathbb{S}^{d-1}} |a_n^{1/\gamma}[s(p,\hat{\Theta}_n(t)) - s(p,\Theta_I)] - a_n^{1/\gamma}[s(p,\hat{\Theta}_n(t^*_{1-\alpha})) - s(p,\Theta_I)] + \tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t)| \\
&\overset{(1)}{=} \sup_{p\in\mathbb{S}^{d-1}} |\mathcal{Z}_n(p,t) - \mathcal{Z}_n(p,t^*_{1-\alpha}) + c^{\rightarrow}_{1-\alpha}(t) + o_p(1)| \\
&\overset{(2)}{=} \sup_{p\in\mathbb{S}^{d-1}} |\mu(t) - \mu(t^*_{1-\alpha}) - (c^{\rightarrow}_{1-\alpha}(t^*_{1-\alpha}) - c^{\rightarrow}_{1-\alpha}(t)) + o_p(1)| = o_p(1),
\end{aligned} \tag{D.26}$$

where (1) follows from $\tilde{c}^{\rightarrow}_{n,b,1-\alpha}(t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)$ by assumption and $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t) = c^{\rightarrow}_{1-\alpha}(t) + o_p(1)$ by $F^{\rightarrow}(\cdot,t)$ being assumed to be continuous and strictly increasing at $c^{\rightarrow}_{1-\alpha}(t)$ and Lemma 11.2.1 in Lehmann and Romano (2005). In (2), we used the fact that $c^{\rightarrow}_{1-\alpha}(t^*_{1-\alpha}) = 0$. The last equality follows from Lemma D.2.

For (ii), the result immediately follows from Theorem 3.3 (i) and the triangle inequality:

$$d_H\big(\mathcal{C}_{1n}(t), \mathcal{C}_{1n}(t')\big) \le d_H\big(\mathcal{C}_{1n}(t), \hat{\Theta}_n(t^*_{1-\alpha})\big)\big) + d_H\big(\mathcal{C}_{1n}(t'), \hat{\Theta}_n(t^*_{1-\alpha})\big)\big). \tag{D.27}$$

This establishes the claim of the theorem. ■

We use the following two lemmas (Lemmas D.3 and D.4) to show Theorem 3.4.

Lemma D.3. *Let S be a compact subset of a finite dimensional Euclidean space. Let $\mathbf{B} \equiv \mathcal{C}(S)$ be the space of continuous functions on S. For a given $g \in \mathbf{B}$, let $\phi_g : \mathbf{B} \to \mathbb{R}$ be defined pointwise by $\phi_g(x) := \sup_{p\in S} g(p) - x(p)$. Then, for any $x \in \mathbf{B}$, ϕ_g is Hadamard directionally differentiable at x, and its directional derivative $\dot{\phi}_g : \mathbf{B} \to \mathbb{R}$ is given pointwise by*

$$\dot{\phi}_g(y) := \sup_{p\in\Psi(g-x)} -y(p), \tag{D.28}$$

where for each $z \in \mathbf{B}$, $\Psi(z) := \mathrm{argmax}_{p\in S} z(p)$. Furthermore, if $\Psi(g-x)$ is singleton-valued, ϕ_g is Hadamard differentiable at x.

Proof of Lemma D.3. The proof is a modification of Theorem 3.1 in Shapiro (1991). First, we show that $\dot{\phi}_g$ is a continuous functional. Let $\{y_n\}$ be a sequence such that $y_n \to y$ for some $y \in \mathbf{B}$. Note that $\Psi(g-x)$ is nonempty and compact by Theorem 17.31 in Aliprantis and Border (2006). Since $-y_n$ converges uniformly to $-y$ on $\Psi(g-x)$, $\max_{p\in\Psi(g-x)} -y_n(p) \to \max_{p\in\Psi(g-x)} -y(p)$. Since the choice of y was arbitrary, $\dot{\phi}_g$ is continuous at every point.

For each p, let $f_p : \mathbf{B} \to \mathbb{R}$ be defined pointwise by $f_p(x) := g(p) - x(p)$. This is a convex functional on $\mathbf{B}$. Since ϕ_g is a pointwise supremum of a family of convex functionals, it is convex. Let $\mathbf{B}^*$ be the dual space of $\mathbf{B}$. For each p, the subdifferential of f_p at y is defined as $\partial f_p(y) := \{f'_p \in \mathbf{B}^* : f_p(z) \ge f_p(y) + f'_p(z-y), \forall z \in \mathbf{B}\}$. We claim that for every y, $\partial f_p(y) = \{-e_p\}$, where e_p is the evaluation map defined by $e_p(z) = z(p)$ for every $z \in \mathbf{B}$. To prove this claim, first note that $-e_p \in \partial f_p(z)$ is obvious. Now suppose there exists $f'_p \in \partial f_p(y)$ such that $f'_p \neq -e_p$. Then, $f_p(z) \ge f_p(y) + f'_p(z-y)$ implies that $y(p) - z(p) \ge f'_p(z-y)$. Since z can be taken arbitrarily, we must have

$$w(p) \ge f'_p(-w) \text{ for all } w \in \mathbf{B}. \tag{D.29}$$

Furthermore, since $f'_p \neq -e_p$, there exists a $w \in \mathbf{B}$ such that $w(p) > f'_p(-w)$. Let $w' := y - w$. Then, $w'(p) = y(p) - w(p) < y(p) - f'_p(-w) = y(p) + f'_p(w) = y(p) + f'_p(y - w')$, which contradicts (D.29). Therefore, $-e_p$ is the unique element of $\partial f_p(y)$.

Fix $y \in \mathbf{B}$. We note that S is a compact subset of a Hausdorff space and that f_p is continuous for every $p \in S$. Furthermore, for any $p \in S$ and $\{p_n\} \subset S$ such that $p_n \to p$, it follows that $f_{p_n}(y) = y(p_n) \to y(p) = f_p(y)$ by the continuity of y. Therefore $p \mapsto f_p(y)$ is continuous at every y. Now the conditions of Theorem 2.4.18 in Zalinescu

(2002) are satisfied. This implies that the subdifferential of ϕ_g at y takes the form:

$$\partial\phi_g(y) = \overline{co}(\cup_{f_p=\sup_{p\in S} f_p}\partial f_p(y)) = co(\cup_{p\in\Psi(g-x)}\{-e_p\}), \tag{D.30}$$

where $\overline{co}(A)$ and $co(A)$ denote the closed convex hull and convex hull of a set A respectively. Here, the closure is taken with respect to the weak-$*$ topology. In the above expression, we used the fact that the set $C := co(\cup_{p\in\Psi(g-x)}\{-e_p\})$ is closed, which we prove below. Let $\{\tilde{e}_n\}$ be a sequence such that $\tilde{e}_n \in C, \forall n$ and $\tilde{e}_n \to \tilde{e}$ for some $\tilde{e} \in \mathbf{B}^*$. Then, by the convexity of C, we may write $\tilde{e}_n = \lambda_n(-e_{p_n}) + (1-\lambda_n)(-e_{p'_n})$ for some sequence $\{(\lambda_n, p_n, p'_n) \in [0,1]\times\Psi(g-x)^2\}$. Since $[0,1]\times\Psi(g-x)^2$ is compact, for any subsequence of $\{(\lambda_n, p_n, p'_n)\}$, there exists a further subsequence $\{(\lambda_{n_{k_j}}, p_{n_{k_j}}, p'_{n_{k_j}})\}$ such that $(\lambda_{n_{k_j}}, p_{n_{k_j}}, p'_{n_{k_j}}) \to (\lambda^*, p^*, p^{**})$ for some $(\lambda^*, p^*, p^{**}) \in [0,1]\times\Psi(g-x)^2$. For each $y \in \mathbf{B}$, it follows that

$$\begin{aligned}\tilde{e}_{n_{k_j}}(y) &= \lambda_{n_{k_j}}(-y(p_{n_{k_j}})) + (1-\lambda_{n_{k_j}})(-y(p'_{n_{k_j}}))\\ &\to \lambda^*(-y(p^*)) + (1-\lambda^*)(-y(p^{**})) = \lambda^*(-e_{p^*}(y)) + (1-\lambda^*)(-e_{p^{**}}(y)) \in C,\end{aligned} \tag{D.31}$$

where the convergence follows from the continuity of y. Since the choice of the subsequence and y was arbitrary, this ensures that the limit $\tilde{e}$ belongs to C. Hence, C is closed.

By Theorem 23.2 in Rockafellar (1970), the Gateaux directional derivative $\dot{\phi}_g^G : \mathbf{B} \to \mathbb{R}$ of ϕ_g satisfies

$$\dot{\phi}_g^G(y) = \sup_{\phi'_g\in\partial\phi_g}\phi'_g(y) = \sup_{\lambda\in[0,1]}\sup_{p,p'\in\Psi(g-x)}\lambda(-y(p)) + (1-\lambda)(-y(p')). \tag{D.32}$$

Now suppose that $\operatorname{argmax}_{\Psi(g-x)} - y(p) = \{\bar{p}\}$ for some $\bar{p} \in \Psi(g-x)$, then the right hand side of (D.32) is equal to $-y(\bar{p})$. Therefore, in this case $\dot{\phi}_g^G(y) = -y(\bar{p}) = \sup_{p\in\Psi(g-x)} -y(p)$. Similarly if $\operatorname{argmax}_{\Psi(g-x)} - y(p)$ is not a singleton, then again the right hand side of (D.32) is equal to $-y(\bar{p})$ for some $\bar{p} \in \Psi(g-x)$ because $-y(\bar{p}) \geq \lambda(-y(\bar{p})) + (1-\lambda)(-y(p))$ for all $\lambda \in [0,1]$ and p, and equality holds only if p is also in $\operatorname{argmax}_{\Psi(g-x)} - y(p)$. Therefore, it follows again that $\dot{\phi}_g^G(y) = \sup_{p\in\Psi(g-x)} -y(p)$. This establishes that $\dot{\phi}_g$ in (D.28) is the Gateaux directional derivative of ϕ_g.

Now we show that the Gateaux directional derivative is actually the Hadamard directional derivative. For any $x, y \in \mathbf{B}$, if $\|x-y\|_\infty \leq \delta$, then $g(p)-x(p)-\delta \leq g(p)-y(p) \leq g(p)-x(p)+\delta$ uniformly. Therefore, $|\phi_g(x)-\phi_g(y)| \leq \delta$. This ensures that ϕ_g is Lipschitz with Lipschitz constant 1. Let $\{t_n\}$ be a sequence such that $t_n \downarrow 0$. Let K be a compact subset of $\mathbf{B}$. For each $y \in K$, it follows that $h_n(y) := [\phi_g(x+t_n y) - \phi_g(x)]/t_n - \dot{\phi}_g^G(y) = o(1)$ because $\dot{\phi}_g^G$ is the Gateaux directional derivative. Furthermore, for any $y, y' \in K$,

$$\begin{aligned}|h_n(y) - h_n(y')| &= |[\phi_g(x+t_n y) - \phi_g(x+t_n y')]/t_n - \dot{\phi}_g^G(y-y')|\\ &\leq \|x+t_n y - (x+t_n y')\|_\infty/t_n + \|y-y'\|_\infty = 2\|y-y'\|_\infty.\end{aligned} \tag{D.33}$$

Therefore, h_n is also Lipschitz. This implies that the family $\{h_n\}$ is equicontinuous on K. Since $h_n \to 0$ pointwise, this ensures $h_n \to 0$ uniformly over K. Since K was arbitrary, this ensures that $\dot{\phi}_g^G$ is the Hadamard directional derivative of ϕ_g. This completes the proof of the first claim.

If $\Psi(g-x)$ is singleton-valued, then $\dot{\phi}_g(az+by) = -az(p^*) - by(p^*) = a\dot{\phi}_g(z) + b\dot{\phi}_g(y)$ for all $a, b \in \mathbb{R}$ and $z, y \in \mathbf{B}$, where p^* is the unique element of $\Psi(g-x)$. Therefore, the second claim follows. ■

Lemma D.4. *Let $m \in \mathbb{N}$. Let $D \subset \mathbb{R}^m$ be a compact convex set with a nonempty interior. Let K_0 be a nonempty closed convex subset of D and $\{\hat{K}_n\}$ be a sequence of measurable closed convex subsets of D. Given a positive sequence $\{\tau_n\}$ such that $\tau_n \to \infty$, let $\mathcal{W}_n := \tau_n(s(\cdot, \hat{K}_n) - s(\cdot, K_0))$. Given $x_0 \in \partial K_0$, let*

$$S^{\to}_{n,x_0} := \tau_n \sup_{p\in\mathbb{S}^{m-1}}\{\langle p, x_0\rangle - s(p, \hat{K}_n)\}_+ \tag{D.34}$$

and

$$L_0 := \underset{p \in \mathbb{S}^{m-1}}{\arg\max} \langle p, x_0 \rangle - s(p, K_0). \tag{D.35}$$

Suppose that $\mathcal{W}_n$ converges weakly to a tight random element $\mathcal{W}$ as $n \to \infty$. Then,

$$S^{\rightarrow}_{n,x_0} \xrightarrow{d} \sup_{p \in L_0} \{-\mathcal{W}(p)\}_+. \tag{D.36}$$

Proof of Lemma D.4. We first note that $x_0 \in \partial K_0$ implies $\sup_{p \in \mathbb{S}^{m-1}} \langle p, x_0 \rangle - s(p, K_0) = 0$. Let $\phi_{x_0} : \mathcal{C}(\mathbb{S}^{m-1}) \to \mathbb{R}$ be defined pointwise by $\phi_{x_0}(f) := \sup_{p \in \mathbb{S}^{m-1}} \langle p, x_0 \rangle - f(p)$. Now the statistic can be written as

$$S^{\rightarrow}_{n,x_0} = \max\{\tau_n(\phi_{x_0}(s(\cdot, \hat{K}_n)) - \phi_{x_0}(s(\cdot, K_0)), 0\}. \tag{D.37}$$

By Lemma D.3, ϕ_{x_0} is Hadamard directionally differentiable at $s(\cdot, K_0)$ with Hadamard directional derivative $\dot{\phi}_{x_0}(y) = \sup_{p \in L_0} -y(p)$. This and the assumption that $\mathcal{W}_n \xrightarrow{u.d.} \mathcal{W}$ ensure the conditions of Theorem 2.1 in Shapiro (1991). It follows that

$$\tau_n(\phi_{x_0}(s(\cdot, \hat{K}_n)) - \phi_{x_0}(s(\cdot, K_0)) \xrightarrow{d} \sup_{p \in L_0} -\mathcal{W}(p). \tag{D.38}$$

The conclusion of the Lemma now follows from (D.37), (D.38), and the continuous mapping theorem. ∎

Proof of Theorem 3.4. Let $t \geq 0$. We apply Lemma D.4 with $D = \Theta$, $K_0 = \Theta_I$, $\hat{K}_n = \hat{\Theta}_n(t)$, and $\tau_n = a_n^{1/\gamma}$. Under our hypothesis, Theorem 3.1 holds. The conclusion of Theorem 3.1 ensures that $\mathcal{W}_n = a_n^{1/\gamma}(s(p, \hat{\Theta}_n(t)) - s(p, \Theta_0))$ converges weakly to a tight limit $\mathcal{W} = \mathcal{Z}(\cdot, t)$. By setting $x_0 = \theta_0$ and $L_0 = \Psi_0$, Lemma D.4 then ensures $T^{\rightarrow}_{n,\theta_0} \xrightarrow{d} \sup_{p \in \Psi_0} \{-\mathcal{Z}(p, t)\}_+$. This completes the proof. ∎

Lemma D.5. *Let $m \in \mathbb{N}$. Let $D \subset \mathbb{R}^m$ be a compact convex set with a nonempty interior. Let K_0 be a nonempty closed convex subset of D and $\{\hat{K}_n\}$ be a sequence of measurable closed convex subsets of D such that*

$$d_H(\hat{K}_n, K_0) = O_p(a_n^{-1/\gamma}), \tag{D.39}$$

for some constant $\gamma > 0$ and positive sequence $\{a_n\}$ such that $a_n \to \infty$. Given $x_0 \in D$, let

$$L_0 := \underset{p \in \mathbb{S}^{m-1}}{\arg\max} \langle p, x_0 \rangle - s(p, K_0). \tag{D.40}$$

Given a positive sequence $\{\kappa_n\}$, let

$$\hat{L}_n := \{p \in \mathbb{S}^{m-1} : \langle p, x_0 \rangle - s(p, \hat{K}_n) \geq \sup_{p' \in \mathbb{S}^{m-1}} [\langle p', x_0 \rangle - s(p', \hat{K}_n)] - \kappa_n / a_n^{1/\gamma}\}. \tag{D.41}$$

Suppose $\kappa_n \to \infty$ and $\kappa_n / a_n^{1/\gamma} \to 0$. Then, $d_H(\hat{L}_n, L_0) = o_p(1)$.

Proof of Lemma D.5. We use Theorem 3.1 in Chernozhukov, Hong, and Tamer (2007) to prove the claim. First note that $\mathbb{S}^{m-1}$ is nonempty and compact. Let $\mathcal{Q}$ and $\mathcal{Q}_n$ be defined pointwise by $\mathcal{Q}(p) := [s(p, K_0) - \langle p, x_0 \rangle] - \inf_{p' \in \mathbb{S}^{m-1}}[s(p', K_0) - \langle p', x_0 \rangle]$ and $\mathcal{Q}_n(p) := [s(p, \hat{K}_n) - \langle p, x_0 \rangle] - \inf_{p' \in \mathbb{S}^{m-1}}[s(p', \hat{K}_n) - \langle p', x_0 \rangle]$. We here note that $\inf_{p' \in \mathbb{S}^{m-1}}[s(p', K_0) - \langle p', x_0 \rangle]$ is finite, and $\inf_{p' \in \mathbb{S}^{m-1}}[s(p', \hat{K}_n) - \langle p', x_0 \rangle]$ is finite almost surely due to the continuity of the objective functions (almost surely for the latter) and $\mathbb{S}^{m-1}$ being compact. Note also that L_0 and $\hat{L}_n$ can be equivalently written as

$$L_0 = \underset{p \in \mathbb{S}^{m-1}}{\arg\min}\, \mathcal{Q}(p) = \{p \in \mathbb{S}^{m-1} : \mathcal{Q}(p) = 0\}, \quad \hat{L}_n = \{p \in \mathbb{S}^{m-1} : \mathcal{Q}_n(p) \leq \kappa_n / a_n^{1/\gamma}\}. \tag{D.42}$$

Since $s(p, K_0)$ is continuous by Theorem 1.1.12 in Li, Ogura, and Kreinovich (2002), $\mathcal{Q}$ is continuous. Similarly, since $s(p, \hat{K}_n)$ is continuous in p for each $\omega \in \Omega$ and measurable for each p, $s(p, \hat{K}_n)$ is jointly measurable by Lemma

4.51 in Aliprantis and Border (2006). Furthermore, $\inf_{p' \in \mathbb{S}^{m-1}}[s(p', \hat{K}_n) - \langle p', x_0 \rangle]$ is measurable by Theorem 2.27 (i) in Molchanov (2005). Thus, $\mathcal{Q}_n$ is jointly measurable.

By (D.39) and Theorem 1.1.12 in Li, Ogura, and Kreinovich (2002), $s(\cdot, \hat{K}_n) - s(\cdot, K_0) = o_p(1)$ uniformly. Therefore, for any $\epsilon > 0$, $\sup_{p \in \mathbb{S}^{m-1}} |s(\cdot, \hat{K}_n) - s(\cdot, K_0)| < \epsilon/2$ with probability approaching 1. This implies that, with probability approaching 1,

$$\begin{aligned}\sup_{p \in \mathbb{S}^{m-1}} |\mathcal{Q}_n(p) - \mathcal{Q}(p)| &\leq \sup_{p \in \mathbb{S}^{m-1}} |s(\cdot, \hat{K}_n) - s(\cdot, K_0)| + |\inf_{p' \in \mathbb{S}^{m-1}}[s(p', \hat{K}_n) - \langle p', x_0 \rangle] - \inf_{p' \in \mathbb{S}^{m-1}}[s(p', K_0) - \langle p', x_0 \rangle]| \\ &< \frac{\epsilon}{2} + |\inf_{p' \in \mathbb{S}^{m-1}}[s(p', K_0) + \frac{\epsilon}{2} - \langle p', x_0 \rangle] - \inf_{p' \in \mathbb{S}^{m-1}}[s(p', K_0) - \langle p', x_0 \rangle]| = \epsilon. \end{aligned} \tag{D.43}$$

Thus $\mathcal{Q}_n - \mathcal{Q} = o_p(1)$ uniformly. Furthermore, uniformly over L_0, $\mathcal{Q}_n(p) = [s(p, K_0) + O_p(a_n^{-1/\gamma}) - \langle p, x_0 \rangle] - \inf_{p' \in \mathbb{S}^{m-1}}[s(p', K_0) + O_p(a_n^{-1/\gamma}) - \langle p', x_0 \rangle] = O_p(a_n^{-1/\gamma})$, where the first equality follows from (D.39), and the second equality follows from the construction of L_0. Hence, under our hypothesis, $\kappa_n \geq \sup_{p \in L_0} a_n^{1/\gamma} \mathcal{Q}_n(p)$ with probability approaching 1. Therefore, all required conditions for Theorem 3.1 (1) in Chernozhukov, Hong, and Tamer (2007) are satisfied. This ensures the claim of the lemma. ■

Proof of Corollary 3.2. (i) Let $\hat{F}_n^{\rightarrow}(x, \theta_0, t)$ be the empirical cdf of $T_{n,\theta_0}^{\rightarrow}(t)$. Similarly, let $F^{\rightarrow}(x, \theta_0, t)$ be the cdf of $\sup_{p \in \Psi_0}\{-\mathcal{Z}(p, t)\}_+$. Note that $d_H(\Psi_0, \hat{\Psi}_n) = o_p(1)$ by Theorem 3.1 of CHT and Lemma D.5 applied with $K_0 = \Theta_I$, $\hat{K}_n = \hat{\Theta}_n(t)$, $L_0 = \Psi_0$, and $\hat{L}_n = \hat{\Psi}_n$. Thus, by Theorem 3.2 with $\Upsilon(x) = \{-x\}_+$, $\Psi_0 = \arg\max_p \langle p, \theta_0 \rangle - s(p, \Theta_I)$, and $\hat{\Psi}_n$ as in (3.21), $\hat{F}_n^{\rightarrow}(x, \theta_0, t) - F^{\rightarrow}(x, \theta_0, t) = o_p(1)$ at each continuity point of $F^{\rightarrow}(\cdot, \theta_0, t)$. The rest of the proof is similar to that of Corollary 3.1 and is therefore omitted. ■

Appendix E: Proof of Theorem 4.1 and Corollaries E.1 and 4.1

In this section, we give the proof of Theorem 4.1, Corollaries E.1 and 4.1 and auxiliary lemmas. In what follows, we use the notation introduced in Section 4.1. Recall that $\Pi(\theta) = \nabla_\theta E[m_\theta]$, and $\mathbb{G}(\theta)$ is a vector of Gaussian processes on Θ whose covariance kernel is $K(\theta, \theta') = E[(m_\theta - E[m_\theta])(m_{\theta'} - E[m_{\theta'}])']$. ς is a J-dimensional vector whose j-th component is such that $\varsigma_j(\theta) = 0$ if $E[m_{j,\theta}] = 0$, $\varsigma_j(\theta) = -\infty$ if $E[m_{j,\theta}] < 0$, and $\varsigma_j(\theta) = \infty$ if $E[m_{j,\theta}] > 0$, and $W(\theta)$ is a $J \times J$ positive definite matrix.

Lemma E.1. *Define*

$$\begin{aligned} f_n(\theta, \lambda, x) &\equiv \|W^{1/2}(\theta)\{(\mathbb{G}(\theta) + \Pi(\theta)\lambda + \sqrt{n} E m_\theta) + x\}\|_+^2, \\ g(\theta, \lambda, x) &\equiv \|W^{1/2}(\theta)\{(\mathbb{G}(\theta) + \Pi(\theta)\lambda + \varsigma(\theta)) + x\}\|_+^2. \end{aligned}$$

Then, the following approximation holds.

$$\inf_{R_{n,u,p}} \ell_n(\theta, \lambda) \stackrel{d}{=} \inf_{R_{n,u,p}} f_n(\theta, \lambda, o_p(1)) = \inf_{R_{u,p}} g(\theta, \lambda, o_p(1)). \tag{E.1}$$

Proof of Lemma E.1. The proof is analogous to that of Lemma A.1 in CHT. The first equality in (E.1) follows by arguing as in Step 2 in the proof of Theorem 4.2 in CHT. For the second equality, we take the following three steps.

Step 1: For any $\epsilon > 0$, by Step 1 in the proof of Lemma A.1 in CHT, we have

$$g(\theta, \lambda, -\epsilon) \leq f_n(\theta, \lambda, -\epsilon) \leq \ell_n(\theta, \lambda) \leq f_n(\theta, \lambda, \epsilon),$$

with probability approaching 1. Therefore, for some $\epsilon_n \downarrow 0$, $\inf_{R_{n,u,p}} g(\theta, \lambda, -\epsilon_n) \leq \inf_{R_{n,u,p}} f_n(\theta, \lambda, \epsilon_n)$ with probability approaching 1.

Step 2: For any $\epsilon_n \downarrow 0$ or $\epsilon_n \uparrow 0$, we have $\inf_{R_{n,u,p}} g(\theta, \lambda, \epsilon_n) \geq \inf_{R_{u,p}} g(\theta, \lambda, \epsilon_n)$ by $R_{n,u,p} \subset R_{u,p}$ for all n.

Step 3: The claim of this step is that for some $\epsilon_n' \downarrow 0$,

$$\inf_{R_{n,u,p}} f_n(\theta,\lambda,\epsilon_n') \le \inf_{R_{u,p}} g(\theta,\lambda,\epsilon_n'),\ wp \to 1.$$

The proof is by contradiction. For this, we note that for any $\theta_n \to \theta \in \Theta_I$, it holds that

$$\begin{aligned}\limsup_{n\to\infty} \sqrt{n}E[m_{j,\theta_n}] &\le \varsigma_j(\theta), \quad \text{if } \varsigma_j(\theta)=0\\ &= \varsigma_j(\theta), \quad \text{if } \varsigma_j(\theta) = -\infty.\end{aligned} \tag{E.2}$$

Now suppose that the claim of this step does not hold. Then, there exist a constant $\epsilon > 0$ and a subsequence $\{n'\}$ of $\{n\}$ such that

$$\lim_{n'\to\infty} [f_{n'}(\theta_{n'},\lambda_{n'},\epsilon_{n'}') - \inf_{R_{u,p}} g(\theta,\lambda,\epsilon)] > 0 \tag{E.3}$$

with probability 1. Passing to a further subsequence if necessary, we may let $(\theta_{n'},\lambda_{n'})$ be such that $(\theta_{n'},\lambda_{n'})$ converges to some $(\theta^*,\lambda^*) \in R_{u,p}$. By the definition of f_n and g, the inequality in (E.3) only occurs if $\limsup_{n'\to\infty} \sqrt{n}'E[m_{j,\theta_{n'}}] > \varsigma_j(\theta^*)$ for some j, which is a contradiction to (E.2).

Finally, combining Steps 1-3, the claim of the Lemma follows. ■

Lemma E.2. *Suppose Assumptions 4.1-4.4 hold. Let $\{\delta_n\}$ be a sequence $\{\delta_n \in \mathbb{R}^J\}$ such that $\|\delta_n\| = o_p(1)$. Then, for each $(u,p) \in \mathbb{R}\times\mathbb{S}^{d-1}$ and any $\epsilon > 0$, there exists a compact set $\bar{R}_{u,p} \subset R_{u,p}$ and N_ϵ such that*

$$P(|\inf_{R_{u,p}} g(\theta,\lambda,\delta_n) - \inf_{\bar{R}_{u,p}} g(\theta,\lambda,\delta_n)| \ge \epsilon) \le \epsilon,\ \textit{for all } n \ge N_\epsilon.$$

Proof of Lemma E.2. Let $A_n \equiv \arg\min_{R_{u,p}} g(\theta,\lambda,\delta_n)$. For any $(\theta^*,\lambda^*) \in A_n$, we therefore have

$$\inf_{R_{u,p}} g(\theta,\lambda,\delta_n) = \|W^{1/2}(\theta^*)\{(\mathbb{G}(\theta^*) + \Pi(\theta^*)\lambda^* + \varsigma(\theta^*)) + \delta_n\}\|_+^2. \tag{E.4}$$

Let $\delta_{j,n}$ be the j-th component of δ_n. Since $\varsigma_j(\theta^*) = -\infty$ for all $j \notin \mathcal{J}(\theta^*)$, it follows that

$$\Pi_j(\theta^*)\lambda^* = v_{j,n}, \quad v_{j,n} := -\mathbb{G}_j(\theta^*) - \delta_{j,n} \quad \forall j \in \mathcal{J}(\theta^*). \tag{E.5}$$

Since Assumption 4.4 (iii) ensures the Slater condition for the convex programming problem in (4.6), there exist Karush-Kuhn-Tucker multipliers $\{\eta_j(\theta^*)\}_{j\in\mathcal{J}(\theta^*)}$ such that $p = \sum_{j\in\mathcal{J}(\theta^*)} \eta_j(\theta^*)\Pi_j(\theta^*)'$. This and (E.5) imply that

$$\langle p, \lambda^*\rangle = \sum_{j\in\mathcal{J}(\theta^*(p))} \eta_j(\theta^*)\Pi_j(\theta^*)\lambda^* = \sum_{j\in\mathcal{J}(\theta^*)} \eta_j(\theta^*) v_{j,n}. \tag{E.6}$$

Hence, for any $(\theta^*,\lambda^*) \in A_n$, λ^* is on the hyperplane defined by (E.6). In particular, the minimum norm solution $\lambda^{**} \equiv \sum_{j\in\mathcal{J}(\theta^*)} \eta_j(\theta^*)v_{j,n}p$ is also on this hyperplane, and $(\theta^*,\lambda^{**}) \in A_n$. Note that $\|\lambda^{**}\| = |\sum_{j\in\mathcal{J}(\theta^*)} \eta_j(\theta^*)v_{j,n}| = O_p(1)$ by $\delta_{j,n} = o_p(1)$ and $\mathbb{G}$ being tight by Assumption 4.2. Let $B_M = \{\lambda : \|\lambda\| \le M\}$ with $M > 0$ and let $\bar{R}_{u,p} \equiv H(p,\Theta_I)\times(K_{u,p}\cap B_M)$. Then, by taking M sufficiently large, one may let $P(A_n \cap \bar{R}_{u,p}) \ge P((\theta^*,\lambda^{**}) \in \bar{R}_{u,p}) \ge 1-\epsilon$ for n sufficiently large. This means that the infimum of $g(\theta,\lambda,\delta_n)$ over $R_{u,p}$ is also achieved on $\bar{R}_{u,p}$ with probability approaching 1. Therefore, there exists N_ϵ such that

$$P(|\inf_{R_{u,p}} g(\theta,\lambda,\delta_n) - \inf_{\bar{R}_{u,p}} g(\theta,\lambda,\delta_n)| \ge \epsilon) \le \epsilon,\ \text{for all } n \ge N_\epsilon.$$

This establishes the claim of the lemma. ■

Proof of Theorem 4.1. It is straightforward to show that Assumptions 4.1-4.3 imply Assumptions 2.1-2.3 using the argument in the proof of Theorem 4.2 in CHT. Hence, it is omitted for brevity. We now show Assumption B.1 below. First, by Assumption 4.4 (iii), for any $\theta \in \Theta_I$, we have $K_{u,p} \cap \sqrt{n}(\Theta - \theta) \to K_{u,p}$. Hence, Assumption B.1 (i) holds with $R_{u,p} = H(p,\Theta_I)\times K_{u,p}$.

Assumption B.1 (ii) then follows from the following steps.

Step 1: Let $\{\delta_n\}$ be a sequence $\{\delta_n \in \mathbb{R}^J\}$ such that $\|\delta_n\| = o_p(1)$. The claim of this step is that for compact set $\bar{R}_{u,p}$ and $\epsilon > 0$, we may approximate $\inf_{(\theta,\lambda)\in\bar{R}_{u,p}} g(\theta,\lambda,\delta_n)$ by $\inf_{(\theta,\lambda)\in M(\epsilon)} g(\theta,\lambda,\delta_n)$ where $M(\epsilon)$ is a finite set. Note that

$$\begin{aligned}\inf_{\bar{R}_{u,p}} g(\theta,\lambda,\delta_n) &= \inf_{(\theta,\lambda)\in\bar{R}_{u,p}} \|W^{1/2}(\theta)\{\mathbb{G}(\theta)+\Pi(\theta)\lambda+\varsigma(\theta)+\delta_n\}\|_+^2 \\ &= \min_{\mathcal{J}} \inf_{(\theta,\lambda)\in\bar{R}_{u,p,\mathcal{J}}} \|W_{\mathcal{J}}^{1/2}(\theta)\{\mathbb{G}_{\mathcal{J}}(\theta)+\Pi_{\mathcal{J}}(\theta)\lambda+\delta_{\mathcal{J},n}\}\|_+^2.\end{aligned}$$

By the stochastic equicontinuity of $(\theta,\lambda)\mapsto(\mathbb{G}(\theta),\Pi(\theta)\lambda,W(\theta))$, there is a finite subset $M(\epsilon)$ of $\bar{R}_{u,p}$ such that

$$P(|\inf_{(\theta,\lambda)\in\bar{R}_{u,p}} g(\theta,\lambda,\delta_n) - \inf_{(\theta,\lambda)\in M(\epsilon)} g(\theta,\lambda,\delta_n)| \geq \epsilon) \leq \epsilon$$

for all n sufficiently large.

Step 2: Now let $(u,p)\in\mathbb{R}\times\mathbb{S}^{d-1}$. Then, for some $\{\delta_n\}$ be a sequence $\{\delta_n\in\mathbb{R}^J\}$, we have

$$\begin{aligned}&\liminf_{n\to\infty} P(\inf_{R_{u,p}} \ell_n(\theta,\lambda) > t) \overset{(i)}{=} \liminf_{n\to\infty} P(\inf_{R_{u,p}} g(\theta,\lambda,\delta_n) > t) \overset{(ii)}{\geq} \liminf_{n\to\infty} P(\inf_{\bar{R}_{u,p}} g(\theta,\lambda,\delta_n) > t+\epsilon/2) - \epsilon/2 \\ &\overset{(iii)}{\geq} \liminf_{n\to\infty} P(\inf_{M(\epsilon)} g(\theta,\lambda,\delta_n) > t+\epsilon) - \epsilon \overset{(iv)}{\geq} P(\inf_{M(\epsilon)} \ell_\infty(\theta,\lambda) > t+\epsilon) - \epsilon \overset{(v)}{\geq} P(\inf_{R_{u,p}} \ell_\infty(\theta,\lambda) > t+\epsilon) - \epsilon, \quad \text{(E.7)}\end{aligned}$$

where (i) follows from Lemma E.1, (ii) follows from Lemma E.2, (iii) follows from Step 2 (finite-dimensional approximability), (iv) follows from the fact that $g(\theta,\lambda,\delta_n)$ converges to ℓ_∞ in finite dimension, and (v) follows from $M(\epsilon)\subset\bar{R}_{u,p}\subset R_{u,p}$. Since ϵ is arbitrary, we have $\liminf_{n\to\infty} P(\inf_{R_{u,p}}\ell_n > t) \geq P(\inf_{R_{u,p}}\ell_\infty > t)$. Similarly, one may show $\limsup_{n\to\infty} P(\inf_{R^o_{u,p}}\ell_n \geq t) \leq P(\inf_{R^o_{u,p}}\ell_\infty \geq t)$. The joint convergence of $\{\inf_{R_{u,p_j}}\ell_n, m=1,\cdots,M\}$ follows similarly. This establishes Assumption B.1 (ii).

Finally, we may write

$$\ell_\infty(\theta,\lambda) = \|W^{1/2}(\theta)\{\mathbb{G}(\theta)+\Pi(\theta)\lambda+\varsigma(\theta)\}\|_+^2 = \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)\{\mathbb{G}_{\mathcal{J}(\theta)}(\theta)+\Pi_{\mathcal{J}(\theta)}(\theta)\lambda\}\|_+^2,\ (\theta,\lambda)\in\Theta_I\times\mathbb{R}^d. \quad \text{(E.8)}$$

This implies that, for each $\theta\in H(p,\Theta_I)$, the set of directions of recessions (see Rockafellar, 1970, Sec. 27) for $\lambda\mapsto\ell_\infty(\theta,\lambda)$ is $\{\lambda : \Pi_j(\theta)'\lambda\leq 0,\ j\in\mathcal{J}(\theta)\}$. The first order condition to the convex programming problem (4.6) implies that there exists positive constants (KKT multipliers) $\{\eta_j(\theta^*)\}_{j\in\mathcal{J}(\theta^*(p))}$ such that $p=\sum_{j\in\mathcal{J}(\theta)}\eta_j(\theta)\Pi_j(\theta)'$. Hence, $\Pi_j(\theta)'\lambda\leq 0,\ j\in\mathcal{J}(\theta)$ implies $\langle p,\lambda\rangle\leq 0$. This and the compactness of Θ_I imply that the set of directions of recessions is a subset of $\{0\}\times\{\lambda:\langle p,\lambda\rangle\leq 0\}$. Again, by compactness of $H(p,\Theta_I)$, the the set of directions of recessions of $R_{u,p}$ is $\{0\}\times\{\lambda:\langle p,\lambda\rangle\geq 0\}$. By Theorem 27.3 in Rockafellar (1970), ℓ_∞ then achieves its minimum on $R_{u,p}$. Similarly, the set $\{(\theta,\lambda):\theta\in H(p,\Theta_I),\ell_\infty(\theta,\lambda)\leq t,\}$ has the set of directions of recessions $\{0\}\times\{\lambda:\langle p,\lambda\rangle\leq 0\}$ by the compactness of $H(p,\Theta_I)$ and (E.8). On the other hand, the objective function $(\theta,\lambda)\mapsto -\langle p,\lambda\rangle$ has the set of directions of recessions $\mathbb{R}^d\times\{\lambda:\langle p,\lambda\rangle\geq 0\}$. Hence, by Theorem 27.3 in Rockafellar (1970), $\inf_{\theta\in H(p,\Theta_I),\lambda\in\Lambda_{\theta,t}} -\langle p,\lambda\rangle$ is finite and achieves its minimum. Hence, $\sup_{\theta\in H(p,\Theta_I)} s(p,\Lambda_{\theta,t})<\infty$. This establishes Assumption B.1 (iii). ■

Corollary E.1. *Suppose Assumptions 4.1-4.5 hold. Then the limiting process $\mathcal{Z}(\cdot,t)$ can be represented as*

$$\mathcal{Z}(p,t) = \sup_{\theta\in H(p,\Theta_I)} \big\{\|\mathcal{R}(p,\theta)\|t^{1/2} - \langle\mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}\rangle\big\}, \quad \text{(E.9)}$$

where $\mathcal{R}(p,\theta) := W^{-1/2}_{\mathcal{J}(\theta)}\left(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)'\right)^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p$. Furthermore, if the weighting matrix satisfies $W_{\mathcal{J}(\theta)}(\theta) = [\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)']^{-1}$ for any $\theta\in\partial\Theta_I$, the limiting process takes the form $\mathcal{Z}(p,t)=\mu(t)+\mathcal{Z}^*(p)$ with $\mu(t)=t^{1/2}$ and

$$\mathcal{Z}^*(p) = \sup_{\theta\in H(p,\Theta_I)} -\langle[\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)']^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p, \mathbb{G}_{\mathcal{J}(\theta)}(\theta)\rangle.$$

Proof of Corollary E.1. Let $s : \partial\Theta \times \mathbb{R}^d \to \mathbb{R}^{J(\theta)}$ be a vector-valued mapping whose j-th component is $s_j(\theta, \lambda) = 1\{\mathbb{G}_j(\theta) + \langle \Pi_j(\theta), \lambda \rangle > 0\}$. The solution λ^* to the minimization problem (4.7) satisfies the following Karush-Kuhn-Tucker (KKT) conditions with probability 1, with a Lagrange multiplier $\mu > 0$:

$$p = 2\mu \Pi_{\mathcal{J}(\theta)}(\theta)' W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*) \tag{E.10}$$

$$t = \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2. \tag{E.11}$$

We can then solve (E.10) to obtain

$$\big(W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)' W^{1/2}_{\mathcal{J}(\theta)}(\theta)\big)^{-1} W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)p = 2\mu W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*). \tag{E.12}$$

Let $\mathcal{R}(p, \theta)$ be the left hand side of the equation above. Take squared norms both sides to obtain

$$\|\mathcal{R}(p, \theta)\|^2 = |2\mu|^2 \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2 = |2\mu|^2 t,$$

where the second equality follows from (E.11). Hence, we obtain

$$2\mu = \|\mathcal{R}(p, \theta)\| t^{-1/2}. \tag{E.13}$$

Plugging this into (E.12) gives

$$W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*) = \frac{\mathcal{R}(p, \theta)}{\|\mathcal{R}(p, \theta)\|} t^{1/2}. \tag{E.14}$$

Substituting (E.13) and (E.14) into (E.10) yields

$$p = \Pi'_{\mathcal{J}(\theta)} W^{1/2}_{\mathcal{J}(\theta)} \mathcal{R}(p, \theta). \tag{E.15}$$

Now, we can use this result to obtain

$$\begin{aligned}
\mathcal{V}(p, \theta, t) &= \langle p, \lambda^* \rangle \\
&= \left\langle \Pi'_{\mathcal{J}(\theta)} W^{1/2}_{\mathcal{J}(\theta)} \mathcal{R}(p, \theta), \lambda^* \right\rangle \\
&= \left\langle \mathcal{R}(p, \theta), W^{1/2}_{\mathcal{J}(\theta)} \Pi_{\mathcal{J}(\theta)} \lambda^* \right\rangle \\
&= \left\langle \mathcal{R}(p, \theta), W^{1/2}_{\mathcal{J}(\theta)} (\Pi_{\mathcal{J}(\theta)} \lambda^* \circ s(\theta, \lambda^*)) \right\rangle \\
&= \left\langle \mathcal{R}(p, \theta), \frac{\mathcal{R}(p, \theta)}{\|\mathcal{R}(p, \theta)\|} t^{1/2} - W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)} \circ s(\theta, \lambda^*)) \right\rangle \\
&= \|\mathcal{R}(p, \theta)\|\, t^{1/2} - \left\langle \mathcal{R}(p, \theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)} \right\rangle,
\end{aligned} \tag{E.16}$$

where the fourth equality follows from the fact that $\mathcal{R}(p, \theta) = \mathcal{R}(p, \theta) \circ s(\theta, \lambda^*)$ by (E.12), and the fifth equality follows from (E.14). Note that $\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)'$ is invertible by Assumption 4.5. Hence, if $W(\theta)$ satisfies $W_{\mathcal{J}(\theta)}(\theta) = (\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}$ for any $\theta \in \partial\Theta_I$, then

$$\|\mathcal{R}(p, \theta)\|^2 = p' \Pi_{\mathcal{J}(\theta)}(\theta)' (\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1} \Pi_{\mathcal{J}(\theta)}(\theta) p. \tag{E.17}$$

Note that Eq. (E.15) implies that $p'p = p'\Pi_{\mathcal{J}(\theta)}(\theta)'(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p = 1$. Combining the results above establishes $\|\mathcal{R}(p, \theta)\| = 1$. Therefore, the limiting process takes the form $\mathcal{Z}(p, t) := \mu(t) + \mathcal{Z}^*(p)$ with $\mu(t) = t^{1/2}$ and

$$\mathcal{Z}^*(p) = \sup_{\theta \in H(p, \Theta_I)} -\langle \mathcal{R}(p, \theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta) \rangle = \sup_{\theta \in H(p, \Theta_I)} -\langle (\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p, \mathbb{G}_{\mathcal{J}(\theta)}(\theta) \rangle. \tag{E.18}$$

This establishes the claim of the Corollary. ■

For deriving the limiting distribution of CHT's statistic, we require the following regularity conditions.

Assumption E.1 (Local Process Regularity for QLR Statistic)**.** *(i) For any finite sets $U \subset \mathbb{R}$ and $S \subset \mathbb{S}^{d-1}$, $(\sup_{R^-_{u,p}} \ell_n, (u,p) \in U \times S) \xrightarrow{d} (\sup_{R^-_{u,p}} \ell_\infty, (u,p) \in U \times S)$. (ii) For any $0 < \epsilon$, there exists $\delta > 0$ such that*

$$\lim_{n\to\infty} P\Big(\sup_{\|p-q\|<\delta} \Big| \sup_{R^-_{u,p}} \ell_n(\theta,\lambda) - \sup_{R^-_{u,q}} \ell_n(\theta,\lambda) \Big| \geq \epsilon \Big) \leq \epsilon, \tag{E.19}$$

where $R^-_{u,p} := H(p,\Theta_I) \times K^-_{u,p}$.

Assumption E.1 (i) requires that the finite dimensional distribution of the supremum of ℓ_n over a class of sets converges to that of ℓ_∞. This is analogous to weak epiconvergence. We call this version "weak supconvergence" as it is close in spirit to Condition S.2 of CHT. CHT's QLR statistic can be written as

$$\sup_{\theta\in\Theta_I} a_n Q_n(\theta) = \max\left\{ \sup_{\theta\in\partial\Theta_I} a_n Q_n(\theta), \sup_{\theta\in\Theta_I^o} a_n Q_n(\theta) \right\}.$$

As the second term on the right hand side asymptotically vanishes by Assumption 4.3, it suffices to study the first term. Using the local process ℓ_n, define

$$\mathcal{L}_n(p,u) := \sup_{\theta\in H(p,\Theta_I)} \sup_{\lambda\in K^-_{u,p}} \ell_n(\theta,\lambda),$$

where $K^-_{u,p} := \{\lambda \in \mathbb{R}^d : \langle p,\lambda\rangle \leq u\}$. Note that $\sup_{p\in\mathbb{S}^{d-1}} \mathcal{L}_n(p,0) = \sup_{\theta\in\partial\Theta_I} a_n Q_n(\theta)$. We therefore study the asymptotic behavior of the process $\mathcal{L}_n(\cdot,u)$ to study that of the QLR statistic.

Lemma E.3. *Suppose the conditions of Corollary E.1 hold. Suppose Assumption E.1 holds. Then $\mathcal{L}_n(\cdot,u) \xrightarrow{u.d.} \mathcal{L}(\cdot,u)$ for each u, and the process $\mathcal{L}$ can be represented as*

$$\mathcal{L}(p,u) = \sup_{\theta\in H(p,\Theta_I)} \|\mathcal{R}(p,\theta)\|^{-1} \Big(\Big\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\Big\rangle + u\Big)^2_+. \tag{E.20}$$

Proof of Lemma E.3. First, by the hypothesis that ℓ_n weakly supconverges to ℓ_∞, we have $\mathcal{L}_n(\cdot,u) \xrightarrow{f.d.} \mathcal{L}(\cdot,u)$ where

$$\mathcal{L}(p,u) := \sup_{\theta\in H(p,\Theta_I)} \sup_{\lambda\in K^-_{u,p}} \|W^{1/2}(\theta)(\mathbb{G}(\theta) + \Pi(\theta) + \varsigma(\theta))\|^2_+. \tag{E.21}$$

The tightness of $\{\mathcal{L}_n(\cdot,u)\}$ follows from the assumption of the corollary, and these results imply $\mathcal{L}_n(\cdot,u) \xrightarrow{u.d.} \mathcal{L}(\cdot,u)$ for each u.

Now we derive the representation of $\mathcal{L}$ given in the theorem. Below, we fix $p \in \mathbb{S}^{d-1}$ and $\theta \in \partial\Theta_I$. As $\theta \in \partial\Theta_I$, the components of $\mathcal{M}(\theta,\lambda)$ for $j \in \mathcal{J}^c(\theta)$ are irrelevant. To obtain a closed form for $\mathcal{L}$, consider the following optimization problem

$$\begin{aligned} \mathcal{C}(\theta,p,u) := \sup_\lambda &\|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda)\|^2_+ \\ &s.t. \langle p,\lambda\rangle \leq u. \end{aligned} \tag{E.22}$$

Similar to the proof of Corollary E.1, the solution λ^* of the problem above satisfies the following KKT conditions with for some Lagrange multiplier $\nu > 0$ with probability 1:

$$\nu p = 2\Pi_{\mathcal{J}(\theta)}(\theta)' W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*) \tag{E.23}$$

$$\langle p,\lambda^*\rangle = u. \tag{E.24}$$

We can solve (E.23) to obtain

$$\nu\mathcal{R}(p,\theta) = 2W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*). \tag{E.25}$$

Taking squared norms both sides, we obtain

$$\nu^2\|\mathcal{R}(p,\theta)\|^2 = 4\|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*)\|^2 4\mathcal{C}(\theta,p,u). \tag{E.26}$$

Plugging in $\nu = 2\mathcal{C}(\theta,p,u)^{1/2}/\|\mathcal{R}(p,\theta)\|$ back to (E.23), we obtain

$$p = \|\mathcal{R}(p,\theta)\|\mathcal{C}(\theta,p,u)^{-1/2}\Pi_{\mathcal{J}(\theta)}(\theta)'W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*). \tag{E.27}$$

Now, substitute this into (E.24),

$$\begin{aligned}
u &= \|\mathcal{R}(p,\theta)\|\mathcal{C}(\theta,p,u)^{-1/2}\Big\langle \Pi_{\mathcal{J}(\theta)}(\theta)'W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*), \lambda^*\Big\rangle \\
&= \|\mathcal{R}(p,\theta)\|\mathcal{C}(\theta,p,u)^{-1/2}\Big\langle W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\Big\rangle \\
&= \|\mathcal{R}(p,\theta)\|\mathcal{C}(\theta,p,u)^{-1/2}\Big\langle \frac{\nu}{2}\mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\Big\rangle \\
&= \Big\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\Big\rangle,
\end{aligned} \tag{E.28}$$

where the second equality follows from (E.25). Using (E.25) and the result above, the right hand side of (E.26) can be alternatively written as

$$2\nu\Big(\Big\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\Big\rangle + \Big\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\Big\rangle\Big) = 2\nu\Big(\Big\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\Big\rangle + u\Big). \tag{E.29}$$

Therefore, from (E.26), we obtain

$$\nu = 2\|\mathcal{R}(p,\theta)\|^{-1}\Big(\Big\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\Big\rangle + u\Big) = 2\|\mathcal{R}(p,\theta)\|^{-1}\Big(\Big\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\Big\rangle + u\Big)_+, \tag{E.30}$$

where the second equality follows from the fact $\nu > 0$. As $\mathcal{C}(\theta,p,u) = \|\mathcal{R}(p,\theta)\|\nu^2/4$, we have

$$\mathcal{C}(\theta,p,u) = \|\mathcal{R}(p,\theta)\|^{-1}\Big(\Big\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\Big\rangle + u\Big)_+^2. \tag{E.31}$$

Take the supremum over $H(p,\Theta_I)$. The result then follows. ■

Proof of Corollary 4.1. We first analyze the Wald statistic $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}_n(p,t) + t^{1/2}\}_+^2$. By Corollary E.1 and the continuous mapping theorem, we may write its weak limit as

$$\begin{aligned}
\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t) + t^{1/2}\}_+^2 &= \sup_{p\in\mathbb{S}^{d-1}}\Big\{\inf_{\theta\in H(p,\Theta_I)}\langle(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p, \mathbb{G}_{\mathcal{J}(\theta)}(\theta)\rangle\Big\}_+^2 \\
&= \sup_{p\in\mathbb{S}^{d-1}}\Big\langle\big(\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))'\big)^{-1}\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))p, \mathbb{G}_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Big\rangle_+^2 = \mathbf{Z},
\end{aligned} \tag{E.32}$$

where we used $H(p,\Theta_I) = \{\theta_I(p)\}$ to obtain the second equality. For the QLR statistic,

$$\sup_{\theta\in\Theta_I} nQ_n(\theta) \xrightarrow{d} \sup_{p\in\mathbb{S}^{d-1}}\mathcal{L}(p,0) \tag{E.33}$$

by Lemma E.3 and the continuous mapping theorem. By Lemma E.3, this limit can be represented as

$$\begin{aligned}
&\sup_{p\in\mathbb{S}^{d-1}}\sup_{\theta\in H(p,\Theta_I)}\Big(\Big\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\Big\rangle\Big)_+^2 \\
&\quad= \sup_{p\in\mathbb{S}^{d-1}}\Big\langle\big(\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))'\big)^{-1}\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))p, \mathbb{G}_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Big\rangle_+^2 = \mathbf{Z}.
\end{aligned} \tag{E.34}$$

This establishes the first claim.

For the second part, note that $\tau^*_{1-\alpha}$ is the $1-\alpha$ quantile of $\mathbf{Z}$. Therefore, it suffices to show that $t^*_{1-\alpha}$ is also the $1-\alpha$ quantile of $\mathbf{Z}$ under our hypotheses. For this, we can write

$$\begin{aligned} t^*_{1-\alpha} &= \inf\Big\{t : P\Big(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \le 0\Big) \ge 1-\alpha\Big\} \\ &= \inf\Big\{t : P\Big(\sup_{p\in\mathbb{S}^{d-1}}\{-t^{1/2}-\mathcal{Z}^*(p)\}_+ \le 0\Big) \ge 1-\alpha\Big\} \\ &= \inf\Big\{t : P\Big(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}^*(p)\}_+ \le t^{1/2}\Big) \ge 1-\alpha\Big\} \\ &= \inf\Big\{t : P(\mathbf{Z}\le t) \ge 1-\alpha\Big\}, \end{aligned} \tag{E.35}$$

where the third equality follows from the fact that for any $x\ge 0$ and a continuous function f, $\sup_{p\in\mathbb{S}^{d-1}}\{-x+f(p)\}_+\le 0 \Leftrightarrow \sup_p\{f(p)\}_+\le x$, and the last equality follows from (E.32). Therefore, the second claim follows. ■

Appendix F: Monte Carlo Experiments

The Wald confidence set $\mathcal{C}_{\text{Wald}}$ is defined by

$$\mathcal{C}_{\text{Wald}} := \{\theta\in\Theta : d(\theta,\hat\Theta_n(t)) \le \hat c^{\rightarrow}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}\},$$

where $t=\ln(\ln n)^{\frac{1}{2}}$. To construct this confidence set, we must compute the support function of the set estimator $\hat\Theta_n(t)$ and the critical value $\hat c^{\rightarrow}_{n,b,1-\alpha}(t)$. For this, we first solve the following problem for a grid of points $p_h\in\mathbb{S}^{d-1}, h=1,\cdots,H$:

$$\begin{aligned} &\max_\theta \langle p_h,\theta\rangle \\ &s.t. \sqrt{n}\sum_{k=1}^{2K}\hat\sigma^{-1}_{k,n}\big(a_k'\theta - F_k(\hat E_n[m(X_i)])\big)_+ \le t. \end{aligned} \tag{F.1}$$

We set H to 100. The problem above is a linear programming problem that can be solved by common softwares. We use Matlab and a high-speed solver generated by a free software `CVXGEN`.[12] The optimized values then give $s(p_h,\hat\Theta_n(t)), h=1,\cdots,H$. We then generate subsamples of size b using Algorithm 3.1 and also compute $s(p_h,\hat\Theta_{n,b,k}(t)), h=1,\cdots,H$ similarly. The subsampling critical value $\hat c^{\rightarrow}_{n,b,1-\alpha}(t)$ is then obtained as the $1-\alpha$-quantile of $\max_{h=1,\cdots H}\sqrt{b}\{s(p_h,\hat\Theta_{n,b,k}(t)) - s(p_h,\hat\Theta_n(t))\}_+$. The critical value for $\mathcal{C}_{\text{Iter}}$ is computed similarly while updating the initial level using Algorithm 3.2. The coverage is checked by comparing the support function of the identified set to that of the confidence set. Specifically, we interpret Θ_I being covered by $\mathcal{C}_{\text{Wald}}$ when $s(p_h,\Theta_I)\le s(p_h,\hat\Theta_n(t)) + \hat c^{\rightarrow}_{n,b,1-\alpha}(t)/\sqrt{n}$ for all h. The Hausdorff loss is then calculated as $d_H(\mathcal{C}_{\text{Wald}},\Theta_I) = \max_{h=1,\cdots,H}|s(p_h,\Theta_I) - s(p_h,\hat\Theta_n(t)) + \hat c^{\rightarrow}_{n,b,1-\alpha}(t)/\sqrt{n}|$.

CHT's confidence set $\mathcal{C}_{\text{CHT-Sub}}$ with a subsampling critical value is defined as

$$\mathcal{C}_{\text{CHT-Sub}} = \{\theta\in\Theta : \sqrt{n}\sum_{k=1}^{2K}\hat\sigma^{-1}_{k,n}\big(a_k'\theta - F_k(\hat E_n[m(X_i)])\big)_+ \le \hat\tau_{n,b,1-\alpha}\}, \tag{F.2}$$

where $\hat\tau_{n,b,1-\alpha}$ is calculated as follows.[13] First, we obtain the boundary of the initial estimator $\hat\Theta_n(t)$ using (F.1). We then introduce a grid of points (10,000 points) inside $\hat\Theta_n(t)$. For each subsample of size b, we then compute:

$$\Gamma_{n,b} = \begin{cases} \sup_{\theta\in\hat\Theta_n(t)}\sqrt{b}\sum_{k=1}^{2K}\hat\sigma^{-1}_{k,b}\big(a_k'\theta - F_k(\hat E_b[m(X_i)])\big)_+, & \text{if } \hat\Theta_n(t)\neq\emptyset, \\ 0, & \text{if } \hat\Theta_n(t)=\emptyset. \end{cases}$$

[12]CVXGEN generates compiled (mex) solvers for linear and quadratic programs and is available for academic purposes. See details at `http://cvxgen.com/docs/index.html`.

[13]This procedure follows the one in Bugni (2010), which is called Subsampling 2.

The subsampling critical value $\hat{\tau}_{n,b,1-\alpha}$ is then computed as the $1-\alpha$ quantile of $\Gamma_{n,b}$. CHT's confidence set $\mathcal{C}_{\text{CHT-Boot}}$ with a bootstrap critical value is defined as in (F.2) but replacing $\hat{\tau}_{n,b,1-\alpha}$ with a bootstrap critical value $\hat{\tau}^*_{n,1-\alpha}$ computed as follows. First, generate a bootstrap sample $\{X_i^*, i = 1, \cdots, n\}$ from the empirical distribution. Compute the bootstrap sample moments $\hat{E}_n[m(X_i^*)]$ and the weights $\hat{\sigma}_n^*$. Given $\hat{\Theta}_n$, compute

$$\Gamma_n^* = \begin{cases} \sup_{\theta\in\hat{\Theta}_n(t)} \sqrt{n} \sum_{k=1}^{2K} (\hat{\sigma}^*_{k,n})^{-1} \big(\{a_k'\theta - F_k(\hat{E}_n[m(X_i^*)])\} - \{a_k'\theta - F_k(\hat{E}_n[m(X_i)])\}\big)_+ & \\ \qquad \times 1\{|a_k'\theta - F_k(\hat{E}_n[m(X_i)])| \leq \kappa_n/\sqrt{n}\} & \text{if } \hat{\Theta}_n(t) \neq \emptyset, \\ 0, & \text{if } \hat{\Theta}_n(t) = \emptyset., \end{cases}$$

where κ_n is one of the following values: $\ln(\ln(n))^{\frac{1}{2}}, \ln(n)^{\frac{1}{2}}$, and $n^{1/8}$. Γ_n^* therefore differs from $\Gamma_{n,b}$ in the following respects. First, it uses the bootstrapped samples instead of subsamples. Second, it re-centers the bootstrapped sample moment $a_k'\theta - F_k(\hat{E}_n[m(X_i^*)])$ in the criterion function by $a_k'\theta - F_k(\hat{E}_n[m(X_i)])$. Third, the term $1\{|a_k'\theta - F_k(\hat{E}_n[m(X_i)])| \leq \kappa_n/\sqrt{n}\}$ selects the moments that are close to be binding but drops others (see detailed discussions in Bugni, 2010). $\hat{\tau}^*_{n,1-\alpha}$ is then computed as the $1-\alpha$ quantile of Γ_n^*. For both $\mathcal{C}_{\text{CHT-Sub}}$ and $\mathcal{C}_{\text{CHT-Boot}}$, we compute their support functions again using (F.2) replacing t with $\hat{\tau}_{n,b,1-\alpha}$ and $\hat{\tau}^*_{n,1-\alpha}$ respectively. The coverage and the Hausdorff loss are then computed by the same procedure used for the Wald confidence sets.

Table 1: The Coverage Probabilities of Confidence Sets

		$n=1000$				$n=500$		
		$K=5$	$K=9$	$K=15$		$K=5$	$K=9$	$K=15$
$\mathcal{C}_{\text{Wald}}$	$b=100$	0.935	0.949	1.000	$b=75$	0.927	0.954	1.000
	$b=150$	0.927	0.936	0.941	$b=100$	0.916	0.929	1.000
	$b=200$	0.922	0.922	0.910	$b=150$	0.897	0.902	0.904
$\mathcal{C}_{\text{Iter}}$	$b=100$	0.938	0.962	1.000	$b=75$	0.982	0.991	1.000
	$b=150$	0.980	0.982	0.985	$b=100$	0.992	0.991	1.000
	$b=200$	0.994	0.991	0.989	$b=150$	0.998	0.996	0.998
$\mathcal{C}_{\text{CHT-Sub}}$	$b=100$	0.760	0.866	1.000	$b=75$	0.893	0.983	1.000
	$b=150$	0.850	0.879	0.965	$b=100$	0.912	0.970	1.000
	$b=200$	0.877	0.899	0.950	$b=150$	0.932	0.969	0.992
$\mathcal{C}_{\text{CHT-Boot}}$	$\kappa_n=\ln(\ln(n))^{\frac{1}{2}}$	0.990	0.994	0.998	$\kappa_n=\ln(\ln(n))^{\frac{1}{2}}$	0.994	0.999	0.999
	$\kappa_n=\ln(n)^{\frac{1}{2}}$	0.995	0.996	0.998	$\kappa_n=\ln(n)^{\frac{1}{2}}$	0.997	0.999	0.999
	$\kappa_n=n^{\frac{1}{8}}$	0.995	0.996	0.998	$\kappa_n=n^{\frac{1}{8}}$	0.996	0.999	0.999

Table 2: The Median Hausdorff Loss of Confidence Sets

		$n=1000$				$n=500$		
		$K=5$	$K=9$	$K=15$		$K=5$	$K=9$	$K=15$
$\mathcal{C}_{\text{Wald}}$	$b=100$	0.186	0.268	0.986	$b=75$	0.262	0.382	1.841
	$b=150$	0.183	0.259	0.313	$b=100$	0.257	0.353	2.039
	$b=200$	0.180	0.252	0.293	$b=150$	0.247	0.332	0.391
$\mathcal{C}_{\text{Iter}}$	$b=100$	0.187	0.278	1.781	$b=75$	0.321	0.472	1.941
	$b=150$	0.217	0.302	0.377	$b=100$	0.359	0.479	2.658
	$b=200$	0.248	0.330	0.397	$b=150$	0.416	0.523	0.647
$\mathcal{C}_{\text{CHT-Sub}}$	$b=100$	0.196	0.260	0.426	$b=75$	0.321	0.429	0.857
	$b=150$	0.222	0.263	0.290	$b=100$	0.333	0.399	0.640
	$b=200$	0.234	0.270	0.279	$b=150$	0.352	0.397	0.423
$\mathcal{C}_{\text{CHT-Boot}}$	$\kappa_n=\ln(\ln(n))^{1/2}$	0.331	0.340	0.292	$\kappa_n=\ln(\ln(n))^{1/2}$	0.490	0.449	0.386
	$\kappa_n=\ln(n)^{1/2}$	0.351	0.352	0.303	$\kappa_n=\ln(n)^{1/2}$	0.527	0.471	0.406
	$\kappa_n=n^{1/8}$	0.345	0.350	0.301	$\kappa_n=n^{1/8}$	0.514	0.465	0.400

Note: For Wald-type confidence sets ($\mathcal{C}_{\text{Wald}}, \mathcal{C}_{\text{Iter}}$) and CHT's confidence set with a subsampling critical value ($\mathcal{C}_{\text{CHT-Sub}}$), the table reports coverage probabilities and median Hausdorff losses under different subsample sizes. For CHT's confidence set with a bootstrap critical value ($\mathcal{C}_{\text{CHT-Boot}}$), the table reports results under different values of κ_n used for moment selection.