A Dual Approach to Inference for Partially Identified Econometric Models*

Hiroaki Kaido
University of California, San Diego
Email: hkaido@ucsd.edu

Job Market Paper

January 8, 2010

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 their quasi-likelihood ratio (QLR)-type statistic. This paper proposes a tractable way to represent the level-set estimator by its support function. The properly normalized (scaled and centered) support function of the level-set estimator 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 collections and confidence sets for Θ_I and θ_0. We also propose a generic step-up algorithm to choose a tuning parameter, the level of the criterion function. 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 some regularity conditions.

JEL Classification: C12
Keywords: *Partial Identification, Set Estimation, Support Function, Weak Epiconvergence, Step-up Procedure, Asymptotic Equivalence*

*I am deeply grateful to Halbert White for his guidance and constant encouragement. I thank Graham Elliott, Ben Gillen, Patrik Guggenberger, Ivana Komunjer, Andres Santos, and Yixiao Sun, Allan Timmermann, and seminar participants at UCSD for their helpful comments. I also benefited from conversations with Jin Seo Cho, Philip Haile, Joel Horowitz, Charles Manski, Jack Porter, and Max Stinchcombe. I gratefully acknowledge financial support from the Nakajima Foundation and UCSD Department of Economics.

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 there), 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. These authors 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 popular approach is to estimate the boundary of Θ_I directly. This estimate can then be used to conduct inference for Θ_I. This is an attractive alternative if the boundary of the identified set is easily estimable. Much of the literature has studied the case where Θ_I is a closed interval (e.g. Horowitz and Manski, 1998, 2000, Manski, 2003, and Imbens and Manski, 2004). Recent studies extend this approach to the case where Θ_I is a multi-dimensional compact convex set (Beresteanu and Molinari, 2008 (BM) and Bontemps, Magnac, and Maurin, 2008). When Θ_I is compact and convex, its support function provides a tractable representation by summarizing the location of the supporting hyperplanes of Θ_I.

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 exploits the wide applicability of the criterion function approach and the tractability of the support function approach. To the best of our knowledge, this is the first such attempt.

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

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

Tamer, 2004, Magnac and Maurin, 2008) 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 $\{t_n\}$. Collecting all the parameter values at which $Q_n(\theta)$ does not exceed the specified level can be computationally demanding. Our alternative method stores the values $\max_{Q_n(\theta)\leq t_n} \langle p, \theta \rangle$ for different unit vectors p. This yields the support function $s(\cdot, \hat{\Theta}_n)$ of the set estimator. The required computation is straightforward, and one can fully recover the set estimator from its support function. This can result in computational savings that range from modest to dramatic.

Another significant contribution here is a new automated step-up algorithm for selecting the tuning parameter t_n. As explained above, the criterion function approach requires the researcher to choose the level t_n of the criterion function to construct the set estimator (CHT; Bugni, 2009). Our iterative algorithm removes the arbitrariness in the choice of t_n. We relate this to a multiple testing problem. Our algorithm can be interpreted as the reduced form of a step-up procedure that controls the familywise error rate (FWER) of hypotheses that are indexed by compact convex sets. This understanding provides a link to Romano and Shaikh's (2009) recent work on a step-down procedure.

Our approach is particularly well suited to conducting hypothesis tests and constructing confidence collections and confidence sets. For this, 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 measure the distance between sets using the Hausdorff metric common in the literature. This enables us to test the hypothesis that the identified set coincides with a given set, i.e., $H_0 : \Theta_I = \Theta_0$. The test can be inverted to construct a confidence collection that contains the identified set as an element, with some prescribed confidence level. Inference methods for this type of hypothesis are as yet unavailable within CHT's framework.

The normalized support function also lets us test whether the identified set includes a specific set or point. That is, for a given set Θ_0 or point θ_0, we can test $H_0 : \Theta_0 \subseteq \Theta_I$ or $H_0 : \theta_0 \in \Theta_I$. The former test can be inverted to construct another confidence collection, containing each subset of the identified set as an element, with at least some prescribed confidence level. Further, taking the union of the elements of this collection yields a confidence set that covers the identified set. This confidence set is comparable to CHT's confidence sets, constructed by inverting their QLR statistic. Similarly, the test for $\theta_0 \in \Theta_I$ can be inverted to construct a confidence set for each point in the identified set. This set is comparable to those of Imbens and Manski (2004), CHT, Romano and Shaikh (2008), and Andrews and Guggenberger (2009).

The construction of confidence collections and confidence sets by inverting the normalized support function was first proposed by BM for the case where Θ_I is a linear transformation of the Aumann expectation of set-valued random variables. Bontemps, Magnac, and Maurin (2007) 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 by extending these results to the general case where Θ_I is the set of minimizers of a criterion function.

Closely related to our work here is that of BM, who develop an estimation and inference framework based on their set-average estimator, a (Minkowski) average of independent and identically distributed (IID) set-valued random variables. One of BM's key ideas is to embed the space of compact convex sets into a subset of the space of continuous functions (Hörmander, 1955; Beer, 1993). In this paper, we follow a similar approach to study the asymptotic behavior of our set estimator. But instead of using a set-averaging approach, we analyze a version of the sample criterion function using *weak epiconvergence* to derive the asymptotic distribution of the normalized support function of the level-set estimator. Weak epiconvergence is a relatively new concept that characterizes the limit of the infimum of stochastic processes over compact sets and has proven useful for studying the asymptotic behavior of extremum estimators with point identification (Knight, 1999; Chernozhukov and Hong, 2004; and Han and Phillips, 2006). Our analysis shows that weak epiconvergence is ideally suited to study extremum estimators of partially identified models[2].

We apply our theory to econometric models characterized by finitely many moment inequalities. This class has been extensively studied. Recent research in this area includes Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2005), CHT, Fan and Park (2007), Galichon and Henry (2007), BM, Guggenberger, Hahn, and Kim (2008), Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2009), Bugni (2009), Canay (2009), Galichon and Henry (2009), Hahn and Ridder (2009), Moon and Schorfheide (2009), and Yıldız (2009). 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, i.e., the union of all elements in the confidence collection constructed from the Wald statistic, is asymptotically equivalent to CHT's confidence set, a level set whose level is a specific quantile of the QLR statistic.

A special case of this result is the equivalence result previously given by BM. They show that the Wald statistic based on their set-average estimator is asymptotically equivalent to CHT's QLR statistic within the class of (one-dimensional) interval-identified models. Our results show that this can be attributed to: (*i*) the asymptotic equivalence of the Wald statistic and the QLR-statistic within a more general class; and (*ii*) the fact that the set-average estimator coincides with the level-set estimator when Θ_I is a closed interval.

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

[2]To the best of our knowledge, Chernozhukov, Hong, and Tamer (2007) is the first article that adapted the idea of weak epiconvergence to partially identified models. They used a modified version, which is called "weak sup-convergence," to study the asymptotic distribution of their QLR statistic. Here we work directly with weak epiconvergence.

the normalized support function and develop our inference methods in section 3. Section 4 studies moment inequality models and presents the equivalence result. We present Monte Carlo simulation results in section 5 and conclude in section 6. We collect together our mathematical proofs in the mathematical appendix.

2 The CHT Framework and Some Useful Background

In this section, we briefly summarize the framework of CHT and introduce basic notions in the theory of variational analysis and random sets.

2.1 Criterion Function Approach

Our first assumption describes the data generation process and the sample and population criterion functions. For this we require the following definition, where we let $\mathbb{R}_+ := [0, \infty)$ and $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$.

Definition 2.1: *Let $\mathcal{S} \subset \mathbb{R}^d$, $d \in \mathbb{N}$. The function $f : \mathcal{S} \to \bar{\mathbb{R}}_+$ is* proper on $\mathcal{S}$ *if $f(x) < \infty$ for at least one $x \in \mathcal{S}$. If f is proper on $\mathcal{S} = \mathbb{R}^d$, we say f is* proper.

Assumption 2.1: *Let $d \in \mathbb{N}$ and $Q : \mathbb{R}^d \to \bar{\mathbb{R}}_+$ be a Borel measurable function. Let $\Theta \subset \mathbb{R}^d$ be compact and convex, with a nonempty interior. Let $(\Omega, \mathfrak{F}, P)$ be a complete probability space. For $n = 1, 2, ...$, let $Q_n : \Omega \times \mathbb{R}^d \to \bar{\mathbb{R}}_+$ be jointly measurable such that $Q_n(\omega, \cdot)$ is proper on Θ for all $\omega \in F \in \mathfrak{F}$, $P(F) = 1$, and for all $\omega \in \Omega$ and $\theta \notin \Theta$, $Q_n(\omega, \theta) = \infty$.*

The set Θ is the parameter space, which we take here to be of finite dimension. Compactness is a standard assumption on Θ for extremum estimation. Convexity and nonempty interior help us to avoid the "parameters on the boundary problem" for partially identified models[3]. The probability measure P governs the stochastic properties of the data generating process (e.g., independence or dependence, stationarity or heterogeneity). When, as is assumed here, $Q_n(\omega, \cdot)$ is proper on Θ for all $\omega \in F \in \mathfrak{F}$, $P(F) = 1$, we say "Q_n is proper on Θ *a.s.*" For convenience in what follows, we define $Q_n(\omega, \cdot)$ outside of Θ to take the value ∞.

The function Q_n acts as our sample criterion function, for example,

$$Q_n(\omega, \theta) = n^{-1} \sum_{i=1}^{n} q(X_i(\omega), \theta) - \inf_{\theta \in \Theta} \; n^{-1} \sum_{i=1}^{n} q(X_i(\omega), \theta),$$

where $\{X_i : \Omega \to \mathbb{R}\}$ is a sequence of random variables and q is a suitable function, e.g., $q(x, \theta) = (x - \theta)^2$ for scalar x and θ. Observe that the second term ensures that we always have $Q_n(\omega, \theta) \geq 0$. As is common, we may write $Q_n(\theta)$ as a shorthand for $Q_n(\,\cdot\,, \theta)$.

[3] This point is already mentioned by CHT, which we do not pursue in this paper. They provided sufficient conditions to ensure the parameters in the interior case. Our assumption is based on Lemma 4.1 of CHT.

Another common choice for Q_n is that associated with generalized method of moments (GMM) estimation,

$$Q_n(\omega,\theta) = [n^{-1}\sum_{i=1}^{n} m(X_i(\omega),\theta)]'\ \hat{V}_n^{-1}(\omega)\ [n^{-1}\sum_{i=1}^{n} m(X_i(\omega),\theta)] \\ - \inf_{\theta\in\Theta}\ [n^{-1}\sum_{i=1}^{n} m(X_i(\omega),\theta)]'\ \hat{V}_n^{-1}(\omega)\ [n^{-1}\sum_{i=1}^{n} m(X_i(\omega),\theta)],$$

where m is a suitable vector-valued function such that $E[m(X_i,\theta)] = 0$ for one or more values of θ, and $\hat{V}_n$ is an estimator of V, a suitably chosen covariance matrix.

The function Q is the population criterion function. Under assumptions given below, Q_n converges to Q in a suitable sense. The population analog Q will thus inherit certain properties (e.g., properness) from the sample criterion function Q_n. Without loss of generality, we normalize the minimum value of Q to 0, i.e. $\inf_\Theta Q(\theta) = 0$. For example, when $\{X_i\}$ is stationary and the expectations exist, the population analog of the first example above is

$$Q(\theta) = E[q(X_i(\cdot),\theta)] - \inf_{\theta\in\Theta}\ E[q(X_i(\cdot),\theta)].$$

Following Chernozhukov, Hong, and Tamer (2007), we define the identified set as the set of minimizers of Q:

Definition 2.2 (Identified set): *The* identified set Θ_I *satisfies*

$$\Theta_I := \{\theta \in \Theta : Q(\theta) = 0\}. \tag{2.1}$$

There are numerous examples where the identified set can be written as in (2.1). See Manski and Tamer (2002), Bajari, Benkard, and Levin (2007), Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008, 2009), Ciliberto and Tamer (2009), and Kaido and White (2009). Leading examples are the cases where Θ_I is a closed interval in $\mathbb{R}$ or an ellipsoid in $\mathbb{R}^2$. Θ_I is a primary object of interest here. In particular, we are concerned with estimation and inference for Θ_I.

We ensure next that Θ_I is a compact convex set contained in the interior of Θ, $\Theta^o :=$ int(Θ).

Assumption 2.2: *(i) Θ_I is nonempty, closed, and convex; (ii) $\Theta_I \subset \Theta^o$.*

The compactness of Θ and Assumption 2.2 (i) imply the compactness of Θ_I. Assumption 2.2 (ii) removes the trivial case $\Theta_I = \Theta$ and the "parameters on the boundary" case. The latter case is definitely of interest, but to keep a tight focus here, we leave this for analysis elsewhere.

Let $\{a_n\}$ be a sequence of positive constants, and define a stochastic process ζ_n on $\mathbb{R}^d$ by

$$\zeta_n(\theta) := a_n Q_n(\theta), \qquad \theta \in \mathbb{R}^d.$$

The constants a_n normalize the criterion function so that ζ_n converges in distribution to a limit process in an appropriate mode, as we discuss further below. We now define the set estimator of interest here as a level set of ζ_n:

Definition 2.3 (Set estimator): *For sequences* $\{t_n \in \mathbb{R}_+\}$ *and* $\{a_n \in \mathbb{R}_+\}$*, the* set estimator *is*

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

To discuss convergence of $\hat{\Theta}_n(t_n)$ to Θ_I, we require suitable distance measures. For this (here and throughout), let $\mathcal{K}$ be a collection of closed subsets in $\mathbb{R}^d$, and let $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^d$. We measure the distance between sets in $\mathcal{K}$ using the following Hausdorff distances.

Definition 2.4 (Directed Hausdorff distance and Hausdorff metric): *For any* $A, B \in \mathcal{K}$*, the* directed Hausdorff distance *is defined as*

$$\vec{d}_H(A, B) := \sup_{a \in A} d(a, B),$$

where $d(a, B) := \inf_{b \in B} \|b - a\|$ *and* $\vec{d}_H(A, B) := \infty$ *if either* A *or* B *is empty. The* Hausdorff metric *is defined as*

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

The directed Hausdorff distance takes the value 0 when $A \subseteq B$ and a positive value otherwise[4]. This is useful in checking the coverage of the set estimator. For convenience, we refer to either of these as "Hausdorff distance measures."

CHT give a set of conditions (C.1 and C.2 in their paper) sufficient for the consistency of $\hat{\Theta}_n(t_n)$ for Θ_I in the Hausdorff metric and for deriving its convergence rate. Those conditions are general enough to be satisfied by many examples involving moment inequalities and equalities. Following CHT's conditions C.1 and C.2, we assume the following.

Assumption 2.3: *(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*

[4] The directed Hausdorff distance is formally the *lower Hausdorff hemimetric*. A hemimetric d defined on a set $\mathbb{E}$ is a mapping $\mathbb{E} \times \mathbb{E} \to \mathbb{R}$ such that for any $x, y, z \in \mathbb{E}$, (i) $d(x, y) \geq 0$, (ii) $d(x, z) \leq d(x, y) + d(y, z)$, and (iii) $d(x, x) = 0$. In other words, a hemimetric satisfies some properties of a metric, but fails to satisfy symmetry ($d(x, y) = d(y, x)$) and identity ($d(x, y) = 0$ if and only if $x = y$). There is also an upper Hausdorff hemimetric, which corresponds to $\vec{d}_H(B, A)$.

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 $1 - \epsilon$.

Under this assumption, the level-set estimator $\hat{\Theta}_n(t_n)$ is consistent in the Hausdorff metric and has a convergence rate $r_n = (a_n / \max\{1, \kappa_n\})^{1/\gamma}$, when t_n satisfies $t_n \geq \sup_{\Theta_I} a_n Q_n(\theta)$ with probability tending to 1. Such a sequence $\{t_n\}$ of levels can be constructed by setting $t_n = t\kappa_n$, where $t > 0$ and κ_n is a slowly diverging sequence, e.g., $\kappa_n = \log \log n$. Theorem A.1 in the Appendix summarizes CHT's consistency and rate of convergence results for interested readers.

The following condition, CHT's degeneracy condition (C.3), often holds for econometric models that involve finitely many moment inequalities.

Assumption 2.4 (Degeneracy): *(i) There is a sequence of subsets Θ_n of Θ, which could be data dependent (i.e., Effros-measurable functions on Ω), 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) $\epsilon_n = O_p(1/a_n^{1/\gamma})$.*

Under this additional condition, CHT show that it is possible to achieve consistency and an exact polynomial rate of convergence by choosing a constant level $t_n = t \in \mathbb{R}_+$. For later use, we summarize the results below.

Theorem 2.1: *Suppose Assumptions 2.1, 2.2, 2.3 (i), (ii), and 2.4 (i) hold. Then, $d_H(\hat{\Theta}_n(t), \Theta_I) = o_p(1)$. Suppose, in addition, Assumption 2.3 (iii) and 2.4 (ii) hold. Then, $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) = O_p(1)$.*

For models with finitely many moment inequalities, the sample criterion function often vanishes on the set $\Theta_n = \{\theta \in \Theta : n^{-1} \sum_{j=1}^n m_{j,\theta} \leq 0\}$, i.e., the set of parameter values at which sample moment inequalities are satisfied. When Θ_I has nonempty interior, the set of points satisfying the moment inequalities approximate Θ_I at $\sqrt{n}$ rate. In this case, Assumption 2.4 holds with $a_n = n$, $\gamma = 2$, and $\kappa_n = 1$. Section 4 studies this class of econometric models.

To keep a tight focus on the goal of unifying the criterion function and support function approaches, we maintain Assumption 2.4 in the following sections.

2.2 Support Function Approach

We begin by defining notions useful for characterizing compact convex sets: *support function*, *supporting plane,* and *support set*. For this, let $\langle x, y \rangle$ denote the (Euclidean) inner product of two vectors $x, y \in \mathbb{R}^d$. We write $\|p\| = \langle p, p \rangle^{1/2}$.

DEFINITION 2.5 (Support function, supporting plane, and support set): *Let $F \in \mathcal{K}$ and $\mathbb{S}^{d-1} := \{p \in \mathbb{R}^d : \|p\| = 1\}$ be the unit sphere in $\mathbb{R}^d$. The* support function *s of F at $p \in \mathbb{S}^{d-1}$ is defined by*

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

The supporting (hyper)plane *$\mathbb{H}(p, F)$ of F at $p \in \mathbb{S}^{d-1}$ is*

$$\mathbb{H}(p, F) = \{x \in \mathbb{R}^d : \langle p, x \rangle = s(p, F)\}.$$

The support set *$H(p, F)$ of F at $p \in \mathbb{S}^{d-1}$is*

$$H(p, F) = \mathbb{H}(p, F) \cap F.$$

The value of the support function $s(p, F)$ measures the signed distance from the origin of the supporting plane $\mathbb{H}(p, F)$ of the set F with a normal vector p. Figure 1 illustrates this. When the set is strictly convex, its support set $H(p, F)$ for each $p \in \mathbb{S}^{d-1}$ is a singleton.

A maximization problem associated with the support function can be utilized to compute the level-set estimator $\hat{\Theta}_n(t)$. Consider the following problem for a given $p \in \mathbb{S}^{d-1}$ and $t \in \mathbb{R}_+$:

$$\begin{aligned} s(p, \hat{\Theta}_n(t)) := \sup_{\theta \in \Theta} \quad & \langle p, \theta \rangle \\ s.t. \quad & a_n Q_n(\theta) \leq t. \end{aligned} \tag{2.2}$$

When Q_n is convex in a neighborhood of Θ_I, this is a convex programming problem, which is straightforward to solve numerically using standard algorithms. Often, such algorithms find a point $\hat{\theta}_n(p, t)$ in the support set $H(p, \hat{\Theta}_n(t))$ as a solution of the problem (2.2). Therefore, a straightforward algorithm to compute the level set estimator is the following.

ALGORITHM 2.1: *Choose $L \in \mathbb{N}$.*

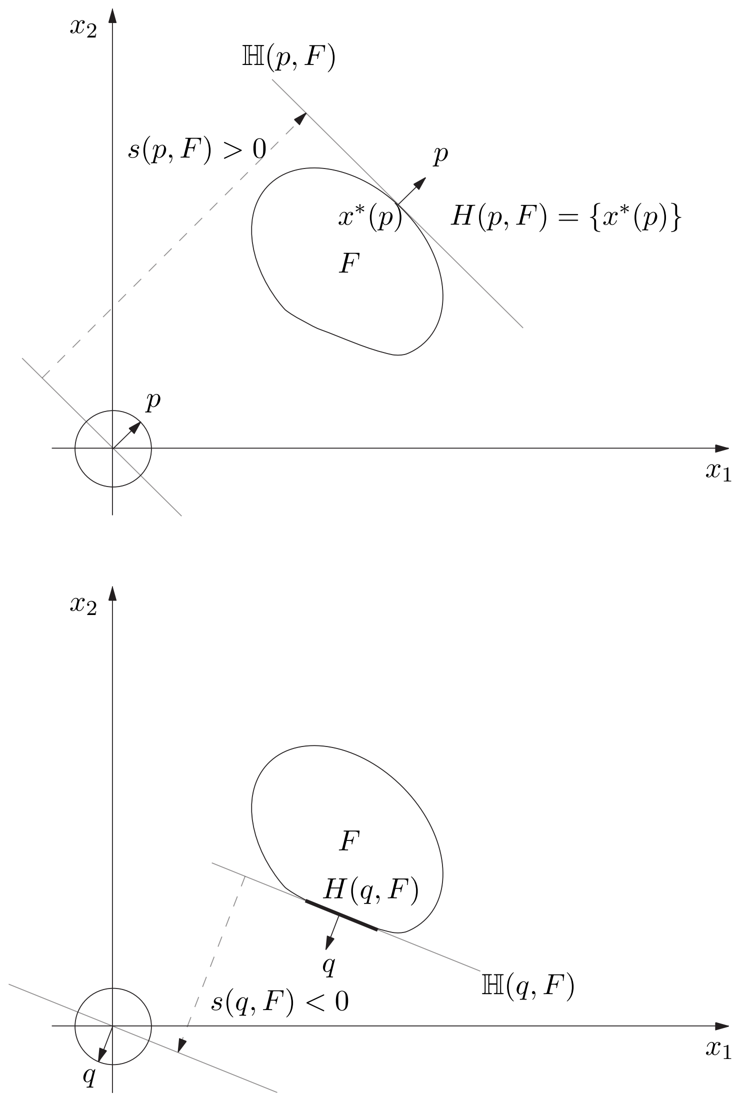
Step 1 : *Generate a grid of points $\{p_1, \cdots, p_L\}$ on the unit sphere $\mathbb{S}^{d-1}$.*

Step 2 : *Solve the problem (2.2) for $p = p_l, l = 1, \cdots, L$. Store the solutions $\{\hat{\theta}_n(p_l, t), l = 1, \cdots, L\}$.*

When the grid is fine enough, the solutions $\{\hat{\theta}_n(p_l, t), l = 1, \cdots, L\}$ provide a good approximation to the boundary of $\hat{\Theta}_n(t)$.

In addition to providing a straightforward algorithm to compute the level set estimator, the support function itself contains useful information. Let $\mathcal{K}_c$ be a collection of compact convex subsets of $\mathbb{R}^d$. Every nonempty compact convex set is the intersection of its supporting

Figure 1: *Support function, supporting plane, and support set*



half spaces. Thus, each element of $\mathcal{K}_c$ is uniquely determined by its support function. This suggests that properties of the metric space $(\mathcal{K}_c, d_H)$ may translate nicely to properties of a space of functions. Let $\mathcal{C}(\mathbb{S}^{d-1})$ be the space of bounded continuous functions on $\mathbb{S}^{d-1}$. Let $\|\cdot\|_{\mathcal{C}(\mathbb{S}^{d-1})}$ be the uniform norm on $\mathbb{S}^{d-1}$: i.e., $\|f\|_{\mathcal{C}(\mathbb{S}^{d-1})} = \sup_{x\in\mathbb{S}^{d-1}} |f(x)|$. Let $d_{\mathcal{C}(\mathbb{S}^{d-1})}$ be the metric induced by this norm. Let $\oplus$ denote the *Minkowski addition* operator, such that $F_1 \oplus F_2 = cl\{f_1 + f_2 : f_1 \in F_1, f_2 \in F_2\}$. The Hörmander embedding theorem is

THEOREM 2.2 (Hörmander's isometric embedding theorem): *The mapping $F \mapsto s(\cdot, F)$ is an isometric embedding of $(\mathcal{K}_c, d_H)$ into a closed convex cone in $(\mathcal{C}(\mathbb{S}^{d-1}), d_{\mathcal{C}(\mathbb{S}^{d-1})})$ that preserves Minkowski addition and non-negative multiplication; i.e. for any F_1 and $F_2 \in \mathcal{K}_c$,*

$$d_H(F_1, F_2) = \|s(\cdot, F_1) - s(\cdot, F_2)\|_{\mathcal{C}(\mathbb{S}^{d-1})} = \sup_{p\in\mathbb{S}^{d-1}} |s(p, F_1) - s(p, F_2)|$$

$$s(p, F_1 \oplus F_2) = s(p, F_1) + s(p, F_2),$$

and for any $\lambda \in \mathbb{R}_+$,

$$s(p, \lambda F_1) = \lambda s(p, F_1).$$

Details for this theorem are in Beer (1993) and Li, Ogura, and Kreinovich (2002)[5].

For our purposes, the fact that the mapping defined by the support function is an isometry is important. Consider the process:

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

This process is useful for conducting inference. Theorem 2.2 ensures that when $\hat{\Theta}_n(t), \Theta_I \in \mathcal{K}_c$, the distance $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I)$ equals $\sup_{p\in\mathbb{S}^{d-1}} |\mathcal{Z}_n(p, t)|$, a functional of $\mathcal{Z}_n(\cdot, t)$.

For the directed Hausdorff distance, we have the following result.

THEOREM 2.3: *Given any two compact convex sets $F_1, F_2 \in \mathcal{K}_c$, the directed Hausdorff distance satisfies*

$$\vec{d}_H(F_1, F_2) = \sup_{p\in\mathbb{B}^d} \{s(p, F_1) - s(p, F_2)\} = \sup_{p\in\mathbb{S}^{d-1}} \{s(p, F_1) - s(p, F_2)\}_+,$$

where $\mathbb{B}^d := \{p \in \mathbb{R}^d : \|p\| \leq 1\}$.

For the proof, see BM Lemma A.1.

[5] Hörmander's embedding theorem holds in a more general environment. If the underlying space $\mathbb{E}$ is separable, then we can isometrically embed $(\mathcal{K}_c(\mathbb{E}), d_H)$ into a closed convex cone in $\mathcal{C}(\mathbb{S}^*)$, where $\mathbb{S}^*$ is the unit sphere in the dual space $\mathbb{E}^*$. We can use metrics that metrize either the strong norm topology or the weak* topology. This permits extending our framework to handle nonparametric estimation, one of our future tasks.

From this result, together with Assumptions 2.1 and 2.2 (i), we have that for given $t \in \mathbb{R}_+$

$$\begin{aligned} a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) &= \sup_{p \in \mathbb{S}^{d-1}} \{-\mathcal{Z}_n(p,t)\}_+ \\ a_n^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \Theta_I) &= \sup_{p \in \mathbb{S}^{d-1}} \{\mathcal{Z}_n(p,t)\}_+. \end{aligned}$$

If for given t we can find a stochastic process $\mathcal{Z}(\cdot, t)$ such that $\mathcal{Z}_n(\cdot, t)$ converges suitably in distribution to $\mathcal{Z}(\cdot, t)$, then the desired limiting distributions of our Hausdorff distance measures follow from the continuous mapping theorem, as these distance measures are continuous functions of $\mathcal{Z}_n(\cdot, t)$. Thus, we focus on deriving the asymptotic distribution of $\mathcal{Z}_n(\cdot, t)$.

As we show, this distribution is a stochastic process on $\mathbb{S}^{d-1}$. In leading cases, this is a Gaussian process. Moreover, its dependence on t is typically straightforward. Specifically, t often affects only the mean of the limiting process and in a manner known a priori. Thus, there exists a known function μ such that for all $t \in \mathbb{R}_+$, $\mathcal{Z}^*(\cdot) := \mathcal{Z}(\cdot, t) - \mu(t)$ is a mean zero process on $\mathbb{S}^{d-1}$, where $\mathcal{Z}(\cdot, t)$ is the desired weak limit of $\mathcal{Z}_n(\cdot, t)$.

2.3 Convergence Concepts

To define the required convergence concepts, consider a sequence of stochastic processes $\{\xi_n\}$ defined on a complete separable metric space $(\mathbb{E}, d)$, so that for $n = 1, 2, ...$, $\xi_n : \Omega \times \mathbb{E} \to \bar{\mathbb{R}}$ is jointly measurable, where $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. For simplicity, we often suppress the dependence of ξ_n on $\omega \in \Omega$, but this should be understood implicitly. In specific contexts, we also view ξ_n as a mapping from the sample space Ω to a space of functions on $\mathbb{E}$.

The simplest convergence in distribution concept for stochastic processes is weak convergence in finite dimensions, defined next. We use the notation $\stackrel{d}{\to}$ to denote the usual convergence in distribution (weak convergence) for a vector of finite dimension (as in, e.g., White, 2001, p.65).

Definition 2.6 (Finite dimensional weak convergence): *Let* $(\mathbb{E}, d)$ *be a complete separable metric space. A sequence of stochastic processes* $\{\xi_n, n \geq 1\}$ *on* $\mathbb{E}$ *is said to* weakly converge in finite dimension *to a limit* ξ*, denoted* $\xi_n \stackrel{f.d.}{\to} \xi$*, if for any finite m-tuple* $(x_1, ..., x_m)$*, where* $x_j \in \mathbb{E}$ *for each* $j = 1, ..., m$*,*

$$(\xi_n(x_1), ..., \xi_n(x_m)) \stackrel{d}{\to} (\xi(x_1), ..., \xi(x_m)).$$

It is well known that the finite dimensional weak convergence is equivalent to weak convergence in the uniform metric when the sequence $\{\xi_n\}$ is *tight* in $l^\infty(\mathbb{E})$, where $l^\infty(\mathbb{E})$ is the space of uniformly bounded functions on $\mathbb{E}$; see, e.g., van der Vaart and Wellner (2000). We denote $\xi_n \stackrel{u.d.}{\to} \xi$ when ξ_n weakly converges to a stochastic process ξ in the uniform metric.

Here, a main goal is to find $\mathcal{Z}(\cdot, t)$ such that $\mathcal{Z}_n(\cdot, t) \stackrel{u.d.}{\to} \mathcal{Z}(\cdot, t)$ for $\mathcal{Z}_n(\cdot, t)$ defined above.

In order to achieve this goal, we make use of the notion of weak epiconvergence given next.

Definition 2.7 (Weak epiconvergence): *A sequence of stochastic processes* $\{\xi_n, n \geq 1\}$ *on* $\mathbb{E}$ *is said to* weakly epiconverge *to a limit* ξ*, denoted* $\xi_n \overset{e.d.}{\to} \xi$*, if for any compact subsets*[6] $R_1, ..., R_k$ *of* $\mathbb{E}$ *with open interiors* $R_1^o, ..., R_m^o$ *and any finite* m*-tuple of real numbers* $\tau_1, ..., \tau_m$*,*

$$
\begin{aligned}
& P\left(\inf_{x\in R_1} \xi(x) > \tau_1, ..., \inf_{x\in R_m} \xi(x) > \tau_m\right) \\
\leq & \liminf_{n\to\infty} P\left(\inf_{x\in R_1} \xi_n(x) > \tau_1, ..., \inf_{x\in R_m} \xi_n(x) > \tau_m\right) \qquad (2.3) \\
\leq & \limsup_{n\to\infty} P\left(\inf_{x\in R_1^o} \xi_n(x) \geq \tau_1, ..., \inf_{x\in R_m^o} \xi_n(x) \geq \tau_m\right) \\
\leq & P\left(\inf_{x\in R_1^o} \xi(x) \geq \tau_1, ..., \inf_{x\in R_m^o} \xi(x) \geq \tau_m\right). \qquad (2.4)
\end{aligned}
$$

We call the condition given by (2.3) the *lower epilimit condition*. Similarly, we call that given by (2.4) the *upper epilimit condition*[7].

Weak epiconvergence is generally useful for studying the limiting distribution of extremum estimators, especially when the criterion function assumes the value infinity, which often occurs in constrained optimization problems[8]. This concept is weaker than weak convergence (on compact sets) in the uniform metric (Pflug, 1995, Proposition 1) and is equivalent to finite dimensional weak convergence when the sequence $\{\xi_n\}$ satisfies a condition called "stochastic equi-lower-semicontinuity" (Knight, 1999, Theorem 2).

For our purposes, weak epiconvergence of a version of the criterion function ζ_n helps ensure the finite dimensional weak convergence of $\mathcal{Z}_n(\cdot, t)$. The desired results then follow by establishing tightness of $\{\mathcal{Z}_n(\cdot, t)\}$.

3 Inference Using the Normalized Support Function

In this section, we present our first main results. We begin by establishing the duality that relates the finite dimensional distribution of the normalized support function $\mathcal{Z}_n(\cdot, t)$ to that of the infimum of a localized criterion function $\tilde{\zeta}_n = a_n Q_n(\theta + \lambda / a_n^{1/\gamma})$ over a class of compact sets. We further show that $\mathcal{Z}_n(\cdot, t)$ converges weakly in the uniform metric to a stochastic process on $\mathbb{S}^{d-1}$ under appropriate regularity conditions on $\tilde{\zeta}_n$. We then present our inference

[6]In this definition, the sets $R_1, ..., R_k$ can instead be taken from a class of relatively compact sets $\mathcal{V}$ such that (i) $\mathcal{V}$ is closed under finite union and intersection; (ii) each compact set K in $\mathbb{E}$ is representable as the intersection of a decreasing sequence in $\mathcal{V}$; and (iii) each open set G in $\mathbb{E}$ is representable as the union of an increasing sequence in $\mathcal{V}$. A typical example for such a $\mathcal{V}$ is a class of closed rectangles. See Pflug (1992) for details.

[7]These names are motivated by Proposition 7.29 in Rockafellar and Wets (2005).

[8]Details on weak epiconvergence can be found in Pflug (1992), Geyer (1994), Pflug (1995), Knight (1999), Geyer (2003), and Molchanov (2005), among others. Recent applications of weak epiconvergence in econometrics include Chernozhukov and Hong (2004), Chernozhukov (2005), and Han and Philips (2006).

methods using functionals of $\mathcal{Z}_n(\cdot,t)$.

3.1 Asymptotic Distribution of the Normalized Support Function

We first add a mild regularity condition on the criterion function. For this, we use the following definition.

Definition 3.1 (Lower semicontinuity): *The function $f:\mathbb{R}^d\to\bar{\mathbb{R}}$ is* lower semicontinuous *(lsc) if $\liminf_{x\to\bar{x}} f(x)\geq f(\bar{x})$ for every $\bar{x}\in\mathbb{R}^d$.*

If a function $f:\Omega\times\mathbb{R}^d\to\bar{\mathbb{R}}$ is such that $f(\omega,\cdot)$ is lsc for all $\omega\in F\in\mathfrak{F}$, $P(F)=1$, then we say f is lower semicontinuous almost surely (lsc $a.s.$). A subtle problem for inference here is that $\hat{\Theta}_n(t)$ may be empty with positive probability in finite samples. To handle this, we use the following convention. We set $s(p,\hat{\Theta}_n(t))=s(p,\hat{\Theta}_n(\underline{t}_n))$ if $\hat{\Theta}_n(t)=\varnothing$, where $\underline{t}_n:=\inf_\Theta a_nQ_n(\theta)$. This convention ensures that $\mathcal{Z}_n(\cdot,t)\in\mathcal{C}(\mathbb{S}^{d-1})$ $a.s.$ Note that $P(\hat{\Theta}_n(t)=\varnothing)\to 0$ under the conditions of Theorem 2.1, so this adjustment becomes less and less likely as $n\to\infty$.

The following lemma establishes the duality between the minimization of the criterion function and the maximization of the corresponding inner product. This lemma provides a way to relate the stochastic behavior of the support function $s(\cdot,\hat{\Theta}_n(t))$ to that of the original criterion function $\zeta_n(\cdot)=a_nQ_n(\cdot)$.

Lemma 3.1 (Duality 1): *Suppose that Assumption 2.1 holds. Let $n\in\mathbb{N}$ and $t\in\mathbb{R}_+$ be given. Suppose ζ_n is lsc a.s. Then, for any $u\in\mathbb{R}$ and $p\in\mathbb{S}^{d-1}$*

$$s(p,\hat{\Theta}_n(t))<u\quad\Leftrightarrow\quad\inf_{\theta\in K_{u,p}\cap\Theta}\zeta_n(\theta)>t,$$

with probability 1, where $K_{u,p}$ is the half space

$$K_{u,p}:=\{\theta\in\mathbb{R}^d:\langle p,\theta\rangle\geq u\}.$$

By this lemma, we can relate the support function of the level set estimator to the criterion function[9]. Our goal is then to relate the normalized support function $\mathcal{Z}_n(p,t)$ to a localized version of the criterion function.

We define a process $\tilde{\zeta}_n$ whose behavior captures that of ζ_n for local deviations from the boundary points of Θ_I. For this, let $\partial\Theta_I$ be the boundary of Θ_I; this coincides with the collection of support points of Θ_I: i.e., $\partial\Theta_I:=\{\theta:\theta\in H(p,\Theta_I),\ p\in\mathbb{S}^{d-1}\}$. Define a stochastic process $\tilde{\zeta}_n$ on $\partial\Theta_I\times\mathbb{R}^d$ by

$$\tilde{\zeta}_n(\theta,\lambda):=\zeta_n(\theta+\lambda/a_n^{1/\gamma}),\qquad\theta\in\partial\Theta_I,\lambda\in\mathbb{R}^d.$$

[9] Note that if $\hat{\Theta}_n(t)=\varnothing$, we take $s(p,\hat{\Theta}_n(t))=\sup_{\theta\in\varnothing}\langle p,\theta\rangle=-\infty$.

The quantity $\theta + \lambda/a_n^{1/\gamma}$ represents a deviation of order $a_n^{-1/\gamma}$ in the direction λ from θ.

To apply the previous lemma, we first note that

$$\{\omega : \mathcal{Z}_n(p,t) < u\} = \left\{\omega : s(p, \hat{\Theta}_n(t)) < s(p, \Theta_I) + u/a_n^{1/\gamma}\right\}.$$

Applying Lemma 3.1 to $s(p, \hat{\Theta}_n(t)) < s(p, \Theta_I) + u/a_n^{1/\gamma}$ yields

$$\{\omega : \mathcal{Z}_n(p,t) < u\} = \left\{\omega : \inf_{\tilde{\theta} \in K_{s(p,\Theta_I)+u/a_n^{1/\gamma},p} \cap \Theta} \zeta_n(\tilde{\theta}) > t\right\}. \tag{3.1}$$

Next, for each $\tilde{\theta} \in K_{s(p,\Theta_I)+u/a_n^{1/\gamma},p} \cap \Theta$, we decompose $\tilde{\theta}$ by letting $\tilde{\theta} = \theta + (\tilde{\theta} - \theta)$ where $\theta \in H(p, \Theta_I)$[10]. We define $\lambda := a_n^{1/\gamma}(\tilde{\theta} - \theta)$. We can therefore write

$$\tilde{\theta} = \theta + \lambda/a_n^{1/\gamma}.$$

The motivation for rescaling λ by $1/a_n^{1/\gamma}$ is that $u/a_n^{1/\gamma}$ appears in the subscript of K in eq. (3.1). By this decomposition, θ represents the part of $\tilde{\theta}$ that gives the inner product value $s(p, \Theta_I)$, and $\lambda/a_n^{1/\gamma}$ represents the part of $\tilde{\theta}$ that gives an inner product value greater than or equal to $u/a_n^{1/\gamma}$. This decomposition is illustrated in figure 2.

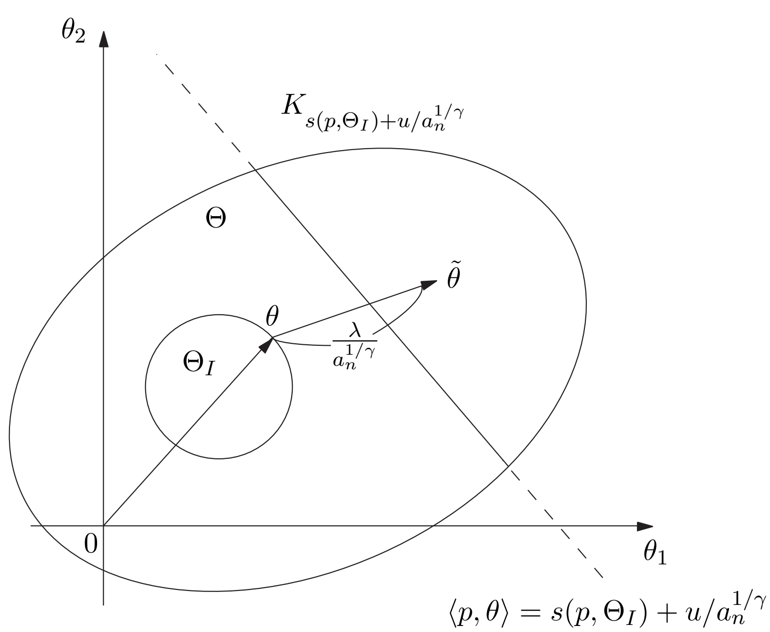


Figure 2: *Decomposition of* $\tilde{\theta}$

[10]Note that θ here is not necessarily unique; however, this has no impact on the arguments to follow.

It is easy to show that λ satisfies $\langle p,\lambda\rangle \geq u$, so $\lambda \in K_{u,p}$. In addition, since $\lambda = a_n^{1/\gamma}(\tilde{\theta}-\theta)$ with $\tilde{\theta} \in \Theta$, λ belongs to a shifted and rescaled space $a_n^{1/\gamma}(\Theta-\theta) := \{\lambda \in \mathbb{R}^d : \lambda = a_n^{1/\gamma}(\tilde{\theta}-\theta), \tilde{\theta} \in \Theta\}$. Thus, $\lambda \in K_{u,p} \cap a_n^{1/\gamma}(\Theta-\theta)$. Using this decomposition, we can rewrite the event in eq. (3.1) as

$$\{\omega : \mathcal{Z}_n(p,t) < u\} = \left\{\omega : \inf_{\theta \in H(p,\Theta_I)} \inf_{\lambda \in K_{u,p} \cap [a_n^{1/\gamma}(\Theta-\theta)]} \tilde{\zeta}_n(\theta,\lambda) > t\right\}.$$

Let $r_{n,u,p}$ be the correspondence defined on $H(p,\Theta_I)$ by

$$r_{n,u,p}(\theta) := K_{u,p} \cap [a_n^{1/\gamma}(\Theta-\theta)], \quad n = 1,2,\ldots\,.$$

For each $\theta \in H(p,\Theta_I)$, the set $K_{u,p} \cap [a_n^{1/\gamma}(\Theta-\theta)]$ is an image of $r_{n,u,p}(\theta)$. Figures 3 and 4 in Appendix B illustrate how this image changes when θ moves along $H(p,\Theta_I)$ for a fixed n[11]. The graph of this correspondence is

$$R_{n,u,p} := \{(\theta,\lambda) : \lambda \in r_{n,u,p}(\theta), \theta \in H(p,\Theta_I)\},$$

which is illustrated in figure 5. Thus,

$$\{\omega : \mathcal{Z}_n(p,t) < u\} = \left\{\omega : \inf_{(\theta,\lambda) \in R_{n,u,p}} \tilde{\zeta}_n(\theta,\lambda) > t\right\}.$$

To analyze this event, it is important to understand the behavior of $R_{n,u,p}$ as n increases.

For this, let $\partial\Theta = \Theta \setminus \Theta^o$ denote the boundary of Θ. We call elements of $\Theta_I \cap \partial\Theta$ *identified parameters on the boundary* (*of* Θ). The remaining elements of Θ_I are *identified parameters in the interior* (*of* Θ). How $R_{n,u,p}$ behaves in the limit depends on whether or not there is an identified parameter on the boundary of Θ.

Specifically, if, as Assumptions 2.1 and 2.2 (*ii*) ensure, there are no identified parameters on the boundary, then $a_n^{1/\gamma}(\Theta-\theta)$ converges to $\mathbb{R}^d$ in the sense of Painlevé-Kuratowski (PK)[12]. Thus, for any (u,p), we have the PK convergences

$$K_{u,p} \cap a_n^{1/\gamma}(\Theta-\theta) \to K_{u,p} \quad \text{and}$$
$$R_{n,u,p} \to R_{u,p} := H(p,\Theta_I) \times K_{u,p}.$$

This case is depicted in figure 6.

[11] Appendix B is available from http://econ.ucsd.edu/~hkaido/pdf/supmat.pdf.

[12] For a sequence $\{C_n\}_{n\in\mathbb{N}}$ of subsets of $\mathbb{R}^d$, the *inner limit* is the set $\liminf_{n\to\infty} C_n := \{x : \exists\{x_n\}_{n\in\mathbb{N}}$ such that $x_n \to x$ and $x_n \in C_n, \forall n\}$ while the *outer limit* is the set $\limsup_{n\to\infty} C_n := \{x : \exists\{x_{n_k}\}_{k\in\mathbb{N}}$ such that $x_{n_k} \to x$ and $x_{n_k} \in C_{n_k}, \forall k\}$. The *limit* of the sequence exists if inner and outer limit sets are equal: $\lim_{n\to\infty} C_n = \liminf_{n\to\infty} C_n = \limsup_{n\to\infty} C_n$. When $\lim_{n\to\infty} C_n$ exists and equal to a set C, the sequence $\{C_n\}_{n\in\mathbb{N}}$ is said to converge to C in the *Painlevé-Kuratowski* sense. See Rockafellar and Wets (2005, ch.4) for details.

On the other hand, if there is an identified parameter on the boundary of Θ, the limit of the sequence of graphs $\{R_{n,u,p}\}_{n=1}^{\infty}$ has a form that depends on the structure of Θ. In this case, the local parameter space may be approximated by a cone, following the ideas of Geyer (1994) and Andrews (1999). This case is definitely of interest, but in order to keep a tight focus here, we leave this for analysis elsewhere.

Our next result provides conditions ensuring that $R_{n,u,p}$ behaves in such a way that the infimum of the stochastic process $\tilde{\zeta}_n$ over $R_{n,u,p}$ is close to the infimum over $R_{u,p}$ in a stochastic sense when n is sufficiently large.

LEMMA 3.2: *Suppose Assumptions 2.1 and 2.2 hold. Suppose that $\tilde{\zeta}_n$ is lsc a.s. and that there exists $\bar{\epsilon} > 0$ such that for any $0 < \epsilon < \bar{\epsilon}$,*

$$\liminf_{n\to\infty} \left\{ (\theta, \lambda) \in R_{u,p} : \tilde{\zeta}_n(\theta, \lambda) < \inf_{R_{u,p}} \tilde{\zeta}_n(\theta, \lambda) + \epsilon \right\} \neq \varnothing, \tag{3.2}$$

almost surely. Then for any $0 < \epsilon < \bar{\epsilon}$ there exists a finite integer N_ϵ such that for all $(u, p) \in \mathbb{R} \times \mathbb{S}^{d-1}$

$$P\left(\left| \inf_{(\theta,\lambda)\in R_{n,u,p}} \tilde{\zeta}_n(\theta, \lambda) - \inf_{(\theta,\lambda)\in R_{u,p}} \tilde{\zeta}_n(\theta, \lambda) \right| \geq \epsilon \right) \leq \epsilon, \quad \forall n \geq N_\epsilon.$$

Since the function $\zeta_n = a_n Q_n$ is defined for all $\theta \in \mathbb{R}^d$, the infima above are well defined. When condition (3.2) holds, we say that $\{\tilde{\zeta}_n\}$ obeys the *nonempty limit ϵ-argmin condition*. This requires that the sequence $\{\tilde{\zeta}_n\}$ stabilizes in such a way that its ϵ-argmin set does not keep moving around. Properness ensures that the difference of the infima in the conclusion is not of the form $\infty - \infty$. This conclusion is an analog of Condition S.1 assumed by CHT, motivated by results of Chernoff (1954) and Andrews (1999).

In order to apply weak epiconvergence to $\tilde{\zeta}_n$, we need to control the limiting behavior of the finite-dimensional distributions of the infima of $\tilde{\zeta}_n$ over a family of compact sets. As $R_{u,p}$ is a closed but unbounded set, we need to replace it with a compact set. As Salinetti and Wets (1986) and Molchanov (2005) show, this can be done under a regularity condition known as *equi-inf-compactness*, defined as follows.

DEFINITION 3.2 (Equi-inf-compactness): *The sequence of stochastic processes $\{\xi_n\}$ is* equi-inf-compact *if for every $\alpha \in \mathbb{R}$ there exists a compact set L_α such that $\{x : \xi_n(x) \leq \alpha\} \subset L_\alpha$ a.s. for all $n \geq 1$.*

If this condition holds for $\{\tilde{\zeta}_n\}$, we can approximate the limit of the infima of $\{\tilde{\zeta}_n\}$ over the closed unbounded set $R_{u,p}$ by the infimum over a compact set $\tilde{R}_{u,p} := R_{u,p} \cap L_{u,p}$ with $L_{u,p}$ properly chosen. Then we can apply weak epiconvergence by checking the limiting behavior of the infima of $\tilde{\zeta}_n$ over compact sets $\{\tilde{R}_{u_j,p_j}, j = 1, 2, ..., m\}$.

We now state a second duality result, relating $\mathcal{Z}_n$ and $\tilde{\zeta}_n$.

Lemma 3.3 (Duality 2): *Suppose that Assumptions 2.1 and 2.2 hold. Let $t \in \mathbb{R}_+$ be given. Suppose that $\{\tilde{\zeta}_n\}$ obeys the nonempty limit ϵ-argmin condition, that $\{\tilde{\zeta}_n\}$ is equi-inf-compact, and that $\tilde{\zeta}_n$ is lsc a.s. for all n sufficiently large. Then, for any finite m-tuple $\{(u_j, p_j) \in \mathbb{R} \times \mathbb{S}^{d-1}\}_{j=1}^m$, there exist compact sets $L_{u_j,p_j}, j = 1, ..., m$, such that, with $\tilde{R}_{u,p} := R_{u,p} \cap L_{u,p}$,*

$$\liminf_{n\to\infty} P(\mathcal{Z}_n(p_1, t) < u_1, ..., \mathcal{Z}_n(p_m, t) < u_m)$$
$$\geq \liminf_{n\to\infty} P\left(\inf_{(\theta,\lambda)\in\tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta, \lambda) > t, ..., \inf_{(\theta,\lambda)\in\tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta, \lambda) > t\right).$$

This lemma ensures that, to study the (finite-dimensional) asymptotic behavior of $\mathcal{Z}_n(\cdot, t)$, it suffices to study the asymptotic behavior of the infima of $\tilde{\zeta}_n$ over compact sets. The right hand side of this inequality can be controlled if $\tilde{\zeta}_n$ weakly epiconverges to a known limiting process $\tilde{\zeta}$. If so, we can seek a process $\mathcal{Z}$ such that

$$P(\mathcal{Z}(p_1, t) < u_1, ..., \mathcal{Z}(p_m, t) < u_m)$$
$$= P\left(\inf_{(\theta,\lambda)\in\tilde{R}_{u_1,p_1}} \tilde{\zeta}(\theta, \lambda) > t, ..., \inf_{(\theta,\lambda)\in\tilde{R}_{u_m,p_m}} \tilde{\zeta}(\theta, \lambda) > t\right).$$

The portmanteu theorem then implies $\mathcal{Z}_n(\cdot, t) \stackrel{f.d.}{\to} \mathcal{Z}(\cdot, t)$.

The next theorem establishes this; it further gives the asymptotic distributions of the Hausdorff distances. For this, we formally impose sufficient regularity on $\{\tilde{\zeta}_n\}$.

Assumption 3.1 (Local Process Regularity): *(i) For all n sufficiently large, $\tilde{\zeta}_n$ is, almost surely, lsc, and Q_n is convex in a neighborhood of Θ_I. (ii) The sequence $\{\tilde{\zeta}_n\}$ obeys the nonempty limit ϵ-argmin condition, is equi-inf-compact, and weakly epiconverges to a stochastic process $\tilde{\zeta}$.*

Assumption 3.1 (ii) is stronger than strictly necessary. It appears that weak epiconvergence can be replaced by the lower epilimit condition without affecting our conclusions.

Theorem 3.1: *Suppose that Assumptions 2.1, 2.2, 2.3, 2.4, and 3.1 hold. For each $t \in \mathbb{R}_+$ and $\theta \in \partial\Theta_I$, let $\hat{\Lambda}(t, \theta)$ be a random level set of the map $\lambda \mapsto \tilde{\zeta}(\theta, \lambda)$ defined by*

$$\hat{\Lambda}(t, \theta) = \{\lambda : \tilde{\zeta}(\theta, \lambda) \leq t\}.$$

Suppose that the limiting process $\tilde{\zeta}$ is such that $\hat{\Lambda}(t, \theta)$ is nonempty a.s. for each $t \in \mathbb{R}_+$ and $\theta \in \partial\Theta_I$.

Then for each $t \in \mathbb{R}_+$,

(i) $\mathcal{Z}_n(\cdot, t) \stackrel{f.d.}{\to} \mathcal{Z}(\cdot, t)$, where $\mathcal{Z}(\cdot, t)$ is a stochastic process on $\mathbb{S}^{d-1}$, which has the repre-

sentation

$$\mathcal{Z}(p,t) = \sup_{\theta \in H(p,\Theta_I)} s\left(p, \hat{\Lambda}(t,\theta)\right); \tag{3.3}$$

(ii) letting m be a finite integer and $\{(u_j, p_j) \in \mathbb{R} \times \mathbb{S}^{d-1}\}_{j=1}^m$ an m-tuple, the finite dimensional distributions of $\mathcal{Z}(\cdot,t)$ satisfy

$$\begin{aligned} & P(\mathcal{Z}(p_1,t) < u_1, ..., \mathcal{Z}(p_m,t) < u_m) \\ & \quad = P\left(\inf_{(\theta,\lambda) \in \tilde{R}_{u_1,p_1}} \tilde{\zeta}(\theta,\lambda) > t, ..., \inf_{(\theta,\lambda) \in \tilde{R}_{u_m,p_m}} \tilde{\zeta}(\theta,\lambda) > t \right); \end{aligned}$$

(iii) $\mathcal{Z}_n(\cdot,t) \overset{u.d.}{\to} \mathcal{Z}(\cdot,t)$, ensuring that

$$\begin{aligned} & a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \overset{d}{\to} \|\mathcal{Z}(\cdot,t)\|_{\mathcal{C}(\mathbb{S}^{d-1})} \\ & a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \overset{d}{\to} \sup_{p \in \mathbb{S}^{d-1}} \{-\mathcal{Z}(p,t)\}_+ \quad \textit{and} \\ & a_n^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \Theta_I) \overset{d}{\to} \sup_{p \in \mathbb{S}^{d-1}} \{\mathcal{Z}(p,t)\}_+ . \end{aligned}$$

3.2 Inference for the Identified Set

Using the asymptotic distribution results of the previous section, we now study hypothesis testing on Θ_I, together with confidence collections and confidence sets constructed by inverting the Wald statistics. To the best of our knowledge, Beresteanu and Molinari (2008) is the first article discussing statistical inference and confidence collections based on the Hausdorff distance measures. Our results below are mostly parallel to the results presented in sections 2.2 and 2.3 of their paper.

We first study hypothesis testing, confidence collections, and confidence sets based on the Hausdorff metric. Let $\Theta_0 \in \mathcal{K}_c$ be a given compact convex set, and consider testing

$$H_0 : \Theta_I = \Theta_0, \quad vs. \quad H_1 : \Theta_I \neq \Theta_0. \tag{3.4}$$

Recall that $d_H(\Theta_I, \Theta_0) = 0$ if and only if $\Theta_I = \Theta_0$. A natural statistic for the test, therefore, is the scaled Hausdorff metric $T_n(t) = a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_0)$. Under the null hypothesis, the statistic has the limiting distribution $\|\mathcal{Z}(\cdot,t)\|_{\mathcal{C}(\mathbb{S}^{d-1})}$ by Theorem 3.1; under the alternative, it diverges to ∞ with probability approaching one because the statistic is not properly centered. Let $\alpha \in (0,1)$ be a significance level. We obtain a test of asymptotic level α by rejecting the null hypothesis when $T_n(t)$ exceeds the asymptotic critical value

$$c_{1-\alpha}(t) := \inf \left\{ x : P\left(\|\mathcal{Z}(\cdot,t)\|_{\mathcal{C}(\mathbb{S}^{d-1})} \leq x \right) \geq 1-\alpha \right\},$$

where $\mathcal{Z}(\cdot,t)$ is the stochastic process given in Theorem 3.1.

It is often difficult to compute this critical value directly, as the required asymptotic distribution differs from case to case. Specifically, the properties of $\mathcal{Z}(\cdot, t)$ depend on the weak epilimit $\tilde{\zeta}$ and therefore on the functional form of the criterion function. Also, the distribution of $\mathcal{Z}(\cdot, t)$ depends on the characteristics of the true identified set Θ_I. For some special cases, it might be possible to simulate the asymptotic distribution of the relevant process to obtain the critical value, but this approach is not generally applicable.

As a practical alternative, we now propose a straightforward subsampling procedure that yields generally valid asymptotic critical values under the high-level assumptions provided above and mild regularity conditions on the rate at which the subsample size grows. For concreteness, we present a procedure for the important class of cases in which the sample criterion function Q_n is constructed from a sample $\{X_i : \Omega \to \mathbb{R}^k\}_{i=1}^n$ of IID random vectors.

ASSUMPTION 3.2: *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,* $k \in \mathbb{N}$.

It is straightforward to extend our results to a sample of stationary and strong mixing time series. See Politis, Romano, and Wolf (1999, Ch 3) for details.

ALGORITHM 3.1 (Subsampling for level-set estimators): *Let* $t > 0$ *and* $0 < \alpha < 1$ *be given. Let* $b := b_n < n$ *be a positive integer. Let* $N_{n,b} = \binom{n}{b}$ *denote the number of subsamples of size b from a sample of size n.*

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* $\zeta_{n,b,k}(X_{k_1}, \cdots, X_{k_b}, \theta) = a_b \tilde{Q}_{n,b,k}(X_{k_1}, \cdots, X_{k_b}, \theta)$, *with the obvious notation.*

Step 2. *For* $k = 1, \cdots, N_{n,b}$, *compute*

$$\hat{T}_{n,b,k}(t) = a_b^{1/\gamma} d_H\left(\hat{\Theta}_{n,b,k}(t), \hat{\Theta}_n(t)\right).$$

Step 3. *Compute the 100×*$(1-\alpha)$*% quantile of the subsampling distribution, given by*

$$\hat{c}_{n,b,1-\alpha}(t) = \inf\left\{x : \hat{F}_{n,b}(x, t) \geq 1 - \alpha\right\},$$

where

$$\hat{F}_{n,b}(x, t) := N_{n,b}^{-1} \sum_{1 \leq k \leq N_{n,b}} 1_{\{\hat{T}_{n,b,k}(t) \leq x\}}.$$

For any t, let $F(x, t) := P[\|\mathcal{Z}(\cdot, t)\|_{\mathcal{C}(\mathbb{S}^{d-1})} \leq x]$ define the cumulative distribution function (CDF) of $\|\mathcal{Z}(\cdot, t)\|_{\mathcal{C}(\mathbb{S}^{d-1})}$. The next theorem is a basic result for subsampling the Hausdorff metric of level set estimators.

Theorem 3.2: *Suppose the conditions of Theorem 3.1 and Assumption 3.2 hold. Suppose that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. Let $\hat{F}_{n,b}(\cdot, t)$ and $\hat{c}_{n,b,1-\alpha}(t)$ be computed by Algorithm 3.1.*

(i) If x is a continuity point of $F(\cdot, t)$, then $\hat{F}_{n,b}(x, t) \to F(x, t)$ in probability;

(ii) If $F(\cdot, t)$ is continuous, then $\sup_x |\hat{F}_{n,b}(x, t) - F(x, t)| \to 0$ in probability;

(iii) If $F(\cdot, t)$ is continuous at $c_{1-\alpha}(t)$, then

$$\lim_{n\to\infty} P\left(a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \hat{c}_{n,b,1-\alpha}(t)\right) = 1 - \alpha.$$

As a corollary of this result, we can show that the test has the correct level and is consistent against any fixed alternative hypothesis.

Corollary 3.1: *Suppose the conditions of Theorem 3.2 hold. Let Θ_0 be a nonempty compact convex subset of Θ^o.*

(i) If $\Theta_0 = \Theta_I$ and $F(\cdot, t)$ is continuous and strictly increasing at $c_{1-\alpha}(t)$, then $\hat{c}_{n,b,1-\alpha}(t) = c_{1-\alpha}(t) + o_p(1)$, and the test has asymptotic rejection probability α:

$$\lim_{n\to\infty} P\left(T_n(t) > \hat{c}_{n,b,1-\alpha}(t)\right) = \alpha.$$

(ii) If $\Theta_0 \neq \Theta_I$, then the test is consistent:

$$\lim_{n\to\infty} P\left(T_n(t) > \hat{c}_{n,b,1-\alpha}(t)\right) = 1.$$

When $N_{n,b}$ is large, we can instead employ a stochastic approximation to $\hat{F}_{n,b}(\cdot, t)$ by randomly drawing subsamples, with or without replacement. See Politis, Romano, and Wolf (1999, Sec. 2.4) for details.

The confidence collection is the collection of compact convex sets such that our test does not reject the null hypothesis in (3.4) when any set in the collection is taken to be Θ_0. This collection can be obtained by inverting our test statistic:

$$\hat{\mathcal{X}}_{n,b,1-\alpha}(t) = \left\{\Psi \in \mathcal{K}_c : a_n^{1/\gamma} d_H\left(\hat{\Theta}_n(t), \Psi\right) \leq \hat{c}_{n,b,1-\alpha}(t)\right\}.$$

We show next that, by the duality of the test, this collection includes Θ_I with probability $1 - \alpha$ asymptotically.

To construct a confidence set, consider the union of the elements of $\hat{\mathcal{X}}_{n,b,1-\alpha}(t)$,

$$\hat{\Psi}_{n,b,1-\alpha}(t) := \bigcup\{\Psi : \Psi \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t)\}.$$

We show that this can be easily computed by expanding our level set estimator by an amount $\hat{\epsilon}_{n,b,1-\alpha}(t) := \hat{c}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}$. This yields a confidence set whose precise level is difficult to determine but which is bounded below by $1 - \alpha$ and is thus conservative.

THEOREM 3.3: *Suppose the conditions of Theorem 3.2 hold. Suppose $F(\cdot, t)$ is continuous at $c_{1-\alpha}(t)$. Then*

(i) $\lim_{n\to\infty} P\left(\Theta_I \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t)\right) = 1-\alpha$;

(ii) $\hat{\Psi}_{n,b,1-\alpha}(t) = \hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}}(t)$, *where* $\hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}}(t)$ *is a closed* $\hat{\epsilon}_{n,b,1-\alpha}(t)$*-envelope of* $\hat{\Theta}_n(t)$ *given by*

$$\hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}}(t) := \{\theta : d_H(\theta, \hat{\Theta}_n(t)) \leq \hat{\epsilon}_{n,b,1-\alpha}(t)\};$$

(iii) Further,

$$\lim_{n\to\infty} P\left(\Theta_I \subseteq \hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}(t)}(t)\right) \geq 1-\alpha.$$

Next, we consider hypothesis testing, confidence collections, and confidence sets based on the directed Hausdorff distance. Again, let $\Theta_0 \in \mathcal{K}_c$ be a given compact convex set. We consider testing

$$H_0 : \Theta_0 \subseteq \Theta_I \quad vs. \quad H_1 : \Theta_0 \nsubseteq \Theta_I. \tag{3.5}$$

Recall that $\vec{d}_H(\Theta_0, \Theta_I) = 0$ if and only if $\Theta_0 \subseteq \Theta_I$. We therefore test this hypothesis using the scaled directed Hausdorff distance $T_n^{\rightarrow}(t) := a_n^{1/\gamma}\vec{d}_H(\Theta_0, \hat{\Theta}_n(t))$. By focusing on the inclusion relationship above, we can directly compare our results to those of BM. We can also test the reverse inclusion using $a_n^{1/\gamma}\vec{d}_H(\hat{\Theta}_n(t), \Theta_0)$ (the scaled upper Hausdorff hemimetric) as described below.

Following BM, we use the triangle inequality

$$\vec{d}_H(\Theta_0, \hat{\Theta}_n(t)) \leq \vec{d}_H(\Theta_0, \Theta_I) + \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)). \tag{3.6}$$

The first term on the right vanishes under the null. Under the assumptions of Theorem 3.1, the second term has a well-defined limiting distribution when scaled by $a_n^{1/\gamma}$. We now specify a subsampling algorithm similar to Algorithm 3.1 to approximate this distribution.

ALGORITHM 3.2: *Implement Algorithm 3.1 but with* $\vec{d}_H(\hat{\Theta}_n(t), \hat{\Theta}_{n,b,k}(t))$ *replacing* $d_H(\hat{\Theta}_{n,b,k}(t), \hat{\Theta}_n(t))$.

The analogs of $\hat{T}_{n,b,k}(t)$, $F_n(x,t)$, $\hat{c}_{n,b,1-\alpha}(t)$, and $\hat{F}_{n,b}(t)$ are denoted $\hat{T}_{n,b,k}^{\rightarrow}(t)$, $F_n^{\rightarrow}(x,t)$, $\hat{c}_{n,b,1-\alpha}^{\rightarrow}(t)$, and $\hat{F}_{n,b}^{\rightarrow}(t)$. Similarly, let $F^{\rightarrow}(x,t) := P(\sup_{p\in\mathbb{S}^{d-1}} \{-\mathcal{Z}(p,t)\}_+ \leq x)$ define the CDF of $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+$. Also, let $c_{1-\alpha}^{\rightarrow}(t) := \inf\{x : P(\sup_{p\in\mathbb{S}^{d-1}} \{-\mathcal{Z}(p,t)\}_+ \leq x) \geq 1-\alpha\}$ define the asymptotic critical value for level $\alpha \in (0,1)$. The next theorem establishes the validity of subsampling for the directed Hausdorff distance.

THEOREM 3.4: *Suppose the conditions of Theorem 3.1 and Assumption 3.2 hold. Suppose that* $b \to \infty$ *and* $b/n \to 0$ *as* $n \to \infty$. *Let* $\hat{F}_{n,b}^{\rightarrow}(\cdot,t)$ *and* $\hat{c}_{n,b,1-\alpha}^{\rightarrow}(t)$ *be computed by Algorithm 3.2.*

(i) If x is a continuity point of $F^{\rightarrow}(\cdot,t)$, *then* $\hat{F}_{n,b}^{\rightarrow}(x,t) \to F^{\rightarrow}(x,t)$ *in probability;*

(ii) If $F^{\rightarrow}(\cdot,t)$ is continuous except at $x=0$, then for any $\epsilon>0$, $\sup_{|x|\geq\epsilon}|\hat{F}^{\rightarrow}_{n,b}(x,t)-F^{\rightarrow}(x,t)|\rightarrow 0$ *in probability;*

(iii) If $F^{\rightarrow}(\cdot,t)$ is continuous at $c^{\rightarrow}_{1-\alpha}(t)$, then

$$\lim_{n\rightarrow\infty} P\left(a_n^{1/\gamma}\vec{d}_H(\Theta_I,\hat{\Theta}_n(t))\leq \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right)=1-\alpha.$$

The directed Hausdorff distance has a discontinuity at $x=0$. The consistency result (ii) is, therefore, weaker than the uniform convergence of subsampling CDFs over the whole real line. It establishes the uniform convergence of the subsampling CDF on compact sets excluding 0. As Bugni (2008) discusses, this weaker consistency result is sufficient for the purpose of hypothesis testing and constructing confidence sets as we are often interested in approximating the $90, 95$, and 99 percentiles. As we will discuss in Section 3.3, it is possible to choose t so that the discontinuity does not occur at the quantiles of interest.

Results for the rejection probability and the consistency against fixed alternatives now follow as before.

COROLLARY 3.2: *Suppose the conditions of Theorem 3.4 hold. Let Θ_0 be a nonempty compact convex subset of Θ^o.*

(i) If $\Theta_0\subseteq\Theta_I$ and $F^{\rightarrow}(\cdot,t)$ is continuous and strictly increasing at $c^{\rightarrow}_{1-\alpha}(t)$, then $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)=c^{\rightarrow}_{1-\alpha}(t)+o_p(1)$, and the test has asymptotic rejection probability bounded above by α:

$$\lim_{n\rightarrow\infty} P\left(T_n^{\rightarrow}(t)>\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right)\leq\alpha;$$

(ii) If $\Theta_0\not\subseteq\Theta_I$, then the test is consistent:

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

To construct a confidence collection, invert the test statistic to obtain

$$\hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t)=\left\{\Psi\in\mathcal{K}_c : a_n^{1/\gamma}\vec{d}_H\left(\Psi,\hat{\Theta}_n(t)\right)\leq\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right\}.$$

As we show, this gives a conservative confidence collection. To construct a confidence set, consider the union of the elements of $\hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t)$,

$$\hat{\Psi}^{\rightarrow}_{n,b,1-\alpha}(t):=\bigcup\{\Psi:\Psi\in\hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t)\}.$$

Analogous to our previous result, we show that this can be easily computed by expanding our level set estimator by an amount $\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t):=\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}$. In contrast to our previous result, this confidence set has asymptotic level $1-\alpha$. This is a Wald-type confidence set that is directly comparable to the QLR-type confidence set studied by CHT.

Theorem 3.5: *Suppose the conditions of Theorem 3.4 hold. Suppose $F^{\rightarrow}(\cdot,t)$ is continuous at $c^{\rightarrow}_{1-\alpha}(t)$. Then*

(i) For each $\Theta_0 \subseteq \Theta_I$, $\lim_{n\to\infty} P\left(\Theta_0 \in \hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t)\right) \geq 1-\alpha$ with equality when $\Theta_0 = \Theta_I$;

(ii) $\hat{\Psi}^{\rightarrow}_{n,b,1-\alpha}(t) = \hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}}(t)$, where $\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}}(t)$ is a closed $\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)$-envelope of $\hat{\Theta}_n(t)$ given by

$$\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}}(t) := \left\{\theta : d(\theta, \hat{\Theta}_n(t)) \leq \hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)\right\};$$

(iii) Furthermore, for t small enough,

$$\lim_{n\to\infty} P\left(\Theta_I \subseteq \hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}}(t)\right) = 1-\alpha.$$

Note that for the confidence set to achieve the coverage probability $1-\alpha$, we must set t small enough. We will discuss how to choose t in the next subsection.

Results for testing the reverse inclusion

$$H_0 : \Theta_I \subseteq \Theta_0 \quad vs. \quad H_1 : \Theta_I \nsubseteq \Theta_0. \tag{3.7}$$

follow similarly, based on the statistic $T_n^{\leftarrow}(t) := a_n^{1/\gamma}\vec{d}_H(\hat{\Theta}_n(t), \Theta_0)$. A subsampling algorithm that can be used to approximate the relevant limiting distribution is

Algorithm 3.3: *Implement Algorithm 3.1 but with $\vec{d}_H(\hat{\Theta}_{n,b,k}(t), \hat{\Theta}_n(t),)$ replacing $d_H(\hat{\Theta}_{n,b,k}(t), \hat{\Theta}_n(t))$.*

The analogs of $\hat{T}_{n,b,k}(t), F_n(x,t), \hat{c}_{n,b,1-\alpha}(t)$, and $\hat{F}_{n,b}(t)$ are denoted $\hat{T}^{\leftarrow}_{n,b,k}(t), F_n^{\leftarrow}(x,t)$, $\hat{c}^{\leftarrow}_{n,b,1-\alpha}(t)$, and $\hat{F}^{\leftarrow}_{n,b}(t)$. Similarly, let $F^{\leftarrow}(x,t) := P(\sup_{p\in\mathbb{S}^{d-1}}\{\mathcal{Z}(p,t)\}_+ \leq x)$ define the CDF of $\sup_{p\in\mathbb{S}^{d-1}}\{\mathcal{Z}(p,t)\}_+$; and let $c^{\leftarrow}_{1-\alpha}(t) := \inf\{x : P(\sup_{p\in\mathbb{S}^{d-1}}\{\mathcal{Z}(p,t)\}_+ \leq x) \geq 1-\alpha\}$ define the asymptotic critical value for level $\alpha \in (0,1)$. Results analogous to Theorem 3.4, Corollary 3.2, and Theorem 3.5 now follow analogously.

3.3 Choice of Level

As we will see in section 4, 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 $\mu(t)$ nor a consistent estimator, it is possible to remove the arbitrariness in the choice of t. In this section, we show that, at least asymptotically, the choice of t does not matter for constructing confidence sets for Θ_I. The construction of the confidence set is based on Theorem 3.5.

For each $\alpha \in (0,1)$, let $t^*_{1-\alpha}$ be the smallest t such that $c^{\rightarrow}_{1-\alpha}(t) = 0$. That is,

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

We will show that, for any $0 \le t < t^*_{1-\alpha}$, confidence sets constructed in the manner of Theorem 3.5 are asymptotically equivalent to each other, in the sense that their difference (in the Hausdorff metric) is of stochastic order smaller than $a_n^{1/\gamma}$. In this sense, the initial choice of t does not matter for constructing the confidence set, given $t < t^*_{1-\alpha}$.

We start with the following lemma that shows $c^{\rightarrow}_{1-\alpha}$ is non-increasing on $[0, t^*_{1-\alpha}]$.

LEMMA 3.4: *Suppose the conditions of Theorem 3.5 are satisfied. Then, for any $0 \le t < t' \le t^*_{1-\alpha}$,*

$$0 = c^{\rightarrow}_{1-\alpha}(t^*_{1-\alpha}) \le c^{\rightarrow}_{1-\alpha}(t') \le c^{\rightarrow}_{1-\alpha}(t) \le c^{\rightarrow}_{1-\alpha}(0).$$

Recall that a confidence set $\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)}(t)$ is an expansion of the level set $\hat{\Theta}_n(t)$ by the amount $\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}$. Lemma 3.4 suggests that if we start with a large t, the amount we need to expand will be smaller, and at $t = t^*_{1-\alpha}$, we do not need to expand the set at all. The following theorem shows that, when the limiting process takes the form $\mathcal{Z}(p,t) = \mu(t) + \mathcal{Z}^*(p)$, this change in the amount of expansion makes all the confidence sets asymptotically equivalent, so that the initial choice of t is not essential as long as $t < t^*_{1-\alpha}$[13].

THEOREM 3.6: *Suppose the conditions of Theorem 3.5 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. Then for each $\alpha \in (0,1)$ and $0 \le t < t^*_{1-\alpha}$,*

$$d_H\left(\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)}(t), \hat{\Theta}_n(t^*_{1-\alpha})\right) = o_p(a_n^{-1/\gamma}).$$

An immediate corollary is the following.

COROLLARY 3.3: *Suppose that the conditions of Theorem 3.6 hold. Then for each $\alpha \in (0,1)$ and for any $0 \le t \le t'^*_{1-\alpha}$,*

$$d_H\left(\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)}(t), \hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t')}(t'))\right) = o_p(a_n^{-1/\gamma}).$$

Theorem 3.6 raises an interesting research question. CHT construct a confidence set $\hat{\Theta}_n(\hat{\tau}_{n,b,1-\alpha})$ such that $\lim_{n\to\infty} P(\Theta_I \subseteq \hat{\Theta}_n(\hat{\tau}_{n,b,1-\alpha})) = 1-\alpha$, 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 $\sup_{\Theta_I} a_n Q_n(\theta)$. If $t^*_{1-\alpha} = \tau^*_{1-\alpha}$ holds, the confidence sets based on the QLR-approach and our

[13] The reason we cannot allow the equality $t = t^*_{1-\alpha}$ is because the subsampling fails to estimate the quantile at which the distribution is discontinuous.

approach are asymptotically equivalent. The question is under what conditions the asymptotic equivalence holds. In section 4, we give a partial answer to this question. For models that involve finitely many moment inequalities, we will provide conditions on the criterion function and weighting matrix that ensure $t^*_{1-\alpha} = \tau^*_{1-\alpha}$.

Based on these results, we propose a generic algorithm to construct the confidence set.

Algorithm 3.4: *(Iterative Algorithm) Set $\kappa > 0$ small. Initialize $l = 1$, and choose t_l small enough.*

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 3.2, 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 $\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t_l)}(t_l)$.*

Step 2. *Update the level by setting $t_{l+1} := \sup_{\theta \in \hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t_l)}(t_l)} a_n Q_n(\theta)$.*

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 3.6 shows, if the limiting process takes the form $\mathcal{Z}(p,t) = \mu(t) + \mathcal{Z}^*(p)$, one can stop at Step 1, as the iteration does not provide any first-order asymptotic improvement, although it may provide higher order refinements.

The iterative algorithm can be related to a multiple testing problem[14]. Romano and Shaikh (2009) is the first article that considered a step-wise procedure to construct a confidence set for Θ_I using CHT's QLR statistic. As we will show below, Algorithm 3.4 can be interpreted as a reduced form of a step-up procedure that controls the family-wise error rate of a multiple testing problem.

Consider the following family of hypotheses:

$$H_{\Theta_0} : \Theta_0 \subseteq \Theta_I, \quad \Theta_0 \in \mathcal{K}_c. \tag{3.8}$$

Each hypothesis in the family is indexed by a compact convex set Θ_0. For this multiple testing problem, we aim to control the following family-wise error rate (FWER):

$$FWER := P\left(\text{reject at least 1 hypothesis } H_{\Theta_0} \text{ s.t. } \Theta_0 \subseteq \Theta_I\right).$$

Below, we consider subcollections of $\mathcal{K}_c$. It is convenient to match each subcollection with a real number t. For this, we introduce a mapping L_n. For any $t \geq 0$, consider a mapping L_n that assigns a subcollection, $L_n(t) = \{\Theta_0 : \Theta_0 \subseteq \hat{\Theta}_n(t)\}$. Given a subcollection $S \subseteq \mathcal{K}_c$,

[14] Details on the multiple testing problems can be found, for example, in Westfall and Young (1993) and Lehmann and Romano (2005, Ch.8).

it is also possible to define a pseudo inverse mapping M_n by $t = M_n(S) := \inf\{t' : \Theta_0 \subseteq \hat{\Theta}_n(t'), \forall \Theta_0 \in S\}$[15].

We consider the following step-up procedure. The procedure starts with an initial subcollection $S_1 \subseteq \mathcal{K}_c$ of hypotheses. In the first step, we look at hypotheses in S_1^c, the collection of sets that are not in S_1. We then find the hypothesis that gives the least significant statistic value and compare this value with a common critical value. If the least significant statistic's value exceeds the common critical value, we reject all hypotheses in S_1^c and stop. Otherwise, we accept the hypotheses that give values below the common critical value and add them to S_1. We call this new collection S_2. In the next step, we test all hypotheses in S_2^c, which are not accepted in the first step. If the hypothesis with the least significant statistic is rejected, then we reject all the hypotheses in S_2^c and stop. Otherwise, we proceed to test the hypotheses that are not accepted in the first and second step. We repeat this until we stop.

Formally, the step-up procedure can be summarized as follows.

Algorithm 3.5: *(Step-up Procedure) Initialize $l = 1$ and t_l. Set $S_l = L_n(t_l)$.*

Step 1. *Construct the set estimator $\hat{\Theta}_n(t_l)$. If for all $\Theta_0 \in S_l^c$, $a_n^{1/\gamma} \vec{d}_H(\Theta_0, \hat{\Theta}_n(t_l)) > \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t_l)$, then reject all hypotheses in S_l^c and stop.*

Step 2. *Otherwise, set $S_{l+1} = \{\Theta_0 \in \mathcal{K}_c : a_n^{1/\gamma} \vec{d}_H(\Theta_0, \hat{\Theta}_n(t_l)) \leq \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t_l)\}$, set $t_{l+1} = M_n(S_{l+1})$, and proceed.*

Step 3. *Repeat Steps 1-2 until the procedure stops.*

For each l, S_l^c represents the family of hypotheses that are not previously accepted. In each step, the procedure compares the least significant test statistic $\inf_{\Theta_0 \in S_l^c} a_n^{1/\gamma} \vec{d}_H(\Theta_0, \hat{\Theta}_n(t_l))$ and a common critical value $\hat{c}_{n,b,1-\alpha}^{\rightarrow}(t_l)$. In each step, if the procedure does not reject all remaining hypotheses, it creates a random collection of compact convex sets S_l. The procedure continues until it rejects all hypotheses that have not been previously accepted. If the procedure stops after the L-th iteration, we can use S_L as a confidence collection. Further, the union of all sets in this collection is a confidence set for Θ_I by Theorem 3.5 (ii).

If $Q_n(\theta)$ is continuous, the updating rule in Step 2 of Algorithm 3.5 can be written as

$$\begin{aligned} t_{l+1} &= M_n(S_{l+1}) \\ &= \inf\{t : \Theta_0 \subseteq \hat{\Theta}_n(t), \quad \forall \Theta_0 \in S_{l+1}\} \\ &= \inf\left\{ t : \sup_{\bigcup_{\Theta_0 \in S_{l+1}} \Theta_0} a_n Q_n(\theta) \leq t \right\} \\ &= \sup_{\hat{\Theta}_n^{\hat{c}_{n,b,1-\alpha}(t_l)/a_n^{1/\gamma}}(t_l)} a_n Q_n(\theta). \end{aligned}$$

[15] Note that $M_n(L_n(t)) = t$ if Q_n is continuous, but in general $L_n(M_n(S)) \supseteq S$.

Thus, the step-up procedure yields the same sequence $\{t_l, l = 1, 2, \cdots\}$ as Algorithm 3.4. We can view the updating rule of Algorithm 3.4 as a reduced form of the step-up procedure.

We now establish that this procedure asymptotically controls the FWER.

THEOREM 3.7 (Control of FWER): *Suppose the conditions of Theorem 3.6 hold. Then the step-up procedure asymptotically controls the FWER in the strong sense, i.e.,*

$$\lim_{n\to\infty} FWER \leq \alpha,$$

for all possible constellations of true and false hypotheses.

Our treatment of the multiple testing problem defined in eq. (3.8) enables us to use random collections of sets as building blocks. Alternatively, one may consider the family $\{H_{\theta_0} : \theta_0 \in \Theta_I\}$ indexed by $\theta_0 \in \Theta$. In this case, one may construct an analogous step-up procedure that starts with a initial set $\tilde{S}_1 \subset \Theta$ and steps up using the scaled directed Hausdorff distance statistic $a_n^{1/\gamma}\vec{d}_H(\{\theta_0\}, \hat{\Theta}_n(t_l))$. It can be shown that for any $l = 1, 2, \cdots$, $\tilde{S}_l$ equals $\bigcup_{S_l} \Theta_0$; this alternative procedure thus yields the same confidence set as before. This alternative approach is more analogous to Romano and Shaikh (2009)'s step-down procedure, as the hypotheses are indexed by $\theta_0 \in \Theta$. Both the step-up and step-down procedures control the family-wise error rate. Comparing the power of the two procedures is an interesting topic for future work.

3.4 Inference for Points in the Identified Set

As Imbens and Manski (2004) discuss, it is often of interest to test hypotheses regarding the true parameter value that generates the data[16]. As the true parameter value cannot be distinguished from any other element of Θ_I, a relevant question would be whether or not a given parameter value θ_0 is observationally equivalent to the data generating parameter value, i.e. $\theta_0 \in \Theta_I$. The scaled directed Hausdorff distance can be used to test this hypothesis. The test can then be inverted to yield a confidence set that asymptotically covers each point in Θ_I with at least a prespecified probability.

Bontemps, Magnac, and Maurin (2007) extend Imbens and Manskis' (2004) results to set-identified linear models. Our results in this section extend them further to the class of problems that can be studied in the extremum estimation framework.

Let $\theta_0 \in \Theta$, and consider testing

$$H_0 : \theta_0 \in \Theta_I \quad vs. \quad H_1 : \theta_0 \notin \Theta_I. \tag{3.9}$$

This can be equivalently stated as $H_0 : \langle p, \theta_0\rangle \leq s(p, \Theta_I), \forall p \in \mathbb{S}^{d-1}$ $vs.$ $H_1 : \langle p, \theta_0\rangle >$

[16]See also Woutersen (2006), Fan and Park (2007), and Stoye (2009) for extensions of Imbens and Manski's (2004) analysis.

$s(p,\Theta_I), \exists p \in \mathbb{S}^{d-1}$. Suppose for the moment that $\theta_0 \in \partial\Theta_I$. In this case, there exists $p_0 \in \mathbb{S}^{d-1}$ such that $\langle p_0, \theta_0\rangle = s(p_0,\Theta_I)$. This p_0 is a maximizer of $\langle p, \theta_0\rangle - s(p,\Theta_I)$. As such a p_0 may not be unique, let $S_{\theta_0} := \{p \in \mathbb{S}^{d-1} : \langle p, \theta_0\rangle = s(p,\Theta_I)\}$.

For each $n \in \mathbb{N}$, let $\hat{p}_n \in \mathbb{S}^{d-1}$ be a vector that maximizes $\langle p, \theta_0\rangle - s(p,\hat{\Theta}_n(t))$. We use the directed Hausdorff distance statistic to test the hypothesis. Given θ_0, we define the statistic

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

Lemmas in the Mathematical Appendix show $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{d} \{-\mathcal{Z}(p_0,t)\}$ under appropriate regularity conditions.

For each $\alpha \in (0,1)$ and $t \in \mathbb{R}_+$, let

$$
c^{\rightarrow}_{1-\alpha}(p,t) := \inf\left\{x : P\left(\{-\mathcal{Z}(p,t)\}_+ \le x\right) \ge 1-\alpha\right\}.
$$

Similar to the inference for Θ_I, we estimate $c^{\rightarrow}_{1-\alpha}(p,t)$ by subsampling. An aspect specific to pointwise inference is that we use the quantile $c^{\rightarrow}_{1-\alpha}(p_0,t)$ evaluated at p_0.

Algorithm 3.6: *Let $t > 0$ and $0 < \alpha < 1$ be given. Let $b := b_n < n$ be a positive integer. Let $N_{n,b} = \binom{n}{b}$ denote the number of subsamples of size b from a sample of size n.*

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 $\zeta_{n,b,k}(X_{k_1},\cdots,X_{k_b},\theta) = a_b\tilde{Q}_{n,b,k}(X_{k_1},\cdots,X_{k_b},\theta)$, with the obvious notation.*

Step 2. *For $k = 1,\cdots,N_{n,b}$, and for each $p \in \mathbb{S}^{d-1}$, compute $\{-\mathcal{Z}_{n,b,k}(p,t)\}_+$, where*

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

Step 3. *For each $p \in \mathbb{S}^{d-1}$, compute the $100\times(1-\alpha)$% quantile of the subsampling distribution, given by*

$$
\hat{c}^{\rightarrow}_{n,b,1-\alpha}(p,t) = \inf\left\{x : \hat{F}^{\rightarrow}_{n,b}(x,p,t) \ge 1-\alpha\right\},
$$

where

$$
\hat{F}^{\rightarrow}_{n,b}(x,p,t) := N_{n,b}^{-1}\sum_{1\le k\le N_{n,b}} 1_{\left\{\{-\mathcal{Z}_{n,b,k}(p,t)\}_+\le x\right\}}.
$$

For each $p \in \mathbb{S}^{d-1}$ and $t \in \mathbb{R}_+$, let $F^{\rightarrow}(x,p,t) := P\left(\{-\mathcal{Z}(p,t)\}_+ \le x\right)$ define the CDF of $\{-\mathcal{Z}(p,t)\}_+$. The next theorem establishes the validity of subsampling.

Theorem 3.8: *Suppose the conditions of Theorem 3.1 and Assumption 3.2 hold. Suppose that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. Let $\hat{F}^{\to}_{n,b}(\cdot, p, t)$ and $\hat{c}^{\to}_{n,b,1-\alpha}(p, t)$ be computed by Algorithm 3.6. Then for given $t \in \mathbb{R}_+$, $\alpha \in (0, 1)$, and each $p \in \mathbb{S}^{d-1}$,*

(i) If x is a continuity point of $F^{\to}(\cdot, p, t)$, then $\hat{F}^{\to}_{n,b}(x, p, t) \to F^{\to}(x, p, t)$ in probability;

(ii) If $F^{\to}(\cdot, p, t)$ is continuous except at $x = 0$, then for any $\epsilon > 0$, $\sup_{|x|\geq\epsilon} |\hat{F}^{\to}_{n,b}(x, p, t) - F^{\to}(x, p, t)| \to 0$ in probability;

(iii) If $F^{\to}(\cdot, p, t)$ is continuous at $c^{\to}_{1-\alpha}(p, t)$, then

$$\lim_{n\to\infty} P\left(\{-\mathcal{Z}_n(p, t)\}_+ \leq \hat{c}^{\to}_{n,b,1-\alpha}(p, t)\right) = 1 - \alpha.$$

As a corollary of this result, we can show that the test has exact size for some $\theta_0 \in \Theta_I$ and is consistent against any fixed alternative hypothesis.

Corollary 3.4: *Suppose the conditions of Theorem 3.8 hold. Suppose the conditions of Lemma A.7 also hold.*

(i) If $\theta_0 \in \Theta_I$ and if for given $t \in \mathbb{R}_+$ and $\alpha \in (0, 1)$, $F^{\to}(\cdot, p_0, t)$ is continuous and strictly increasing at $c^{\to}_{1-\alpha}(p_0, t)$, then $\hat{c}^{\to}_{n,b,1-\alpha}(\hat{p}_n, t) = c^{\to}_{1-\alpha}(p_0, t) + o_p(1)$, and the test has asymptotic rejection probability α :

$$\lim_{n\to\infty} \sup_{\theta_0\in\Theta_I} P\left(T^{\to}_{n,\theta_0}(t) > \hat{c}^{\to}_{n,b,1-\alpha}(\hat{p}_n, t)\right) = \alpha.$$

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

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

The confidence set for θ_0 is obtained by inverting the test. Define

$$\check{\Theta}_{n,b,1-\alpha}(t) := \{\theta_0 \in \Theta : T^{\to}_{n,\theta_0}(t) \leq \hat{c}^{\to}_{n,b,1-\alpha}(\hat{p}_n, t)\}.$$

Note that $\hat{c}^{\to}_{n,b,1-\alpha}$ depends on θ_0 through $\hat{p}_n$. This dependence of the critical value on θ_0 reflects how precisely each boundary point of Θ_I can be estimated. The following theorem shows that this confidence set has the correct coverage probability.

Theorem 3.9: *Suppose the conditions of Theorem 3.8 hold. Then for a given $\alpha \in (0, 1)$ $\lim_{n\to\infty} \inf_{\theta_0\in\Theta_I} P\left(\theta_0 \in \check{\Theta}_{n,b,1-\alpha}(t)\right) = 1 - \alpha$;*

Note the difference between this confidence set and that for the identified set. To construct $\check{\Theta}_{n,b,1-\alpha}(t)$, we use $\hat{c}^{\to}_{n,b,1-\alpha}(\hat{p}_n, t)$ as a critical value, instead of $\hat{c}^{\to}_{n,b,1-\alpha}(t)$. Intuitively, the former takes into account how precisely the support set of Θ_I for the specific normal vector p_0 is estimated. On the other hand, the latter takes into account how precisely the whole boundary of Θ_I is estimated.

Bontemps, Magnac, Maurin (2007) study a confidence set whose coverage is asymptotically valid uniformly over possible values of a nuisance parameter $\Delta = \sup_{p\in\mathbb{S}^{d-1}} s(p,\Theta_I) - s(-p,\Theta_I)$, the maximum length of the identified set. Developing an extension of Theorem 3.9 in this direction is an interesting topic for future work.

Power against Local Alternatives

So far, our discussion has been based on the fixed probability measure $P = P_{\theta_0,\Theta_I}$, where θ_0 is the parameter associated with the true DGP and Θ_I the identified set. In this section, we consider the power of the test against a sequence of alternatives $\{P_{\theta_n,\Theta_I}\}$ indexed by parameter values $\{\theta_n\}$ while fixing the identified set.

If θ_0 is in the interior of Θ_I, it can be shown that the test has no power against alternatives in the neighborhood of θ_0. In the following, we therefore study the case where $\theta_0 \in \partial\Theta_I$.

Let $\pi_{n,b,t} : \Theta \to [0,1]$ be the power function defined by

$$\pi_{n,b,t}(\theta) := P_{\theta,\Theta_I}\left(T^{\rightarrow}_{n,\theta}(t) > c^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n,t)\right).$$

Let $h > 0$. Consider the sequence of alternatives $\{\theta_n\}$ that satisfies $\theta_n := \theta_0 + \lambda/a_n^{1/\gamma}$, where $\theta_0 \in \partial\Theta_I$, and $\lambda \in \mathbb{R}^d$ satisfies $\langle p_0, \lambda\rangle = h$. In terms of the support function, the local parameter has the property that $\langle p_0, \theta_n\rangle = s(p_0,\Theta_I) + h/a_n^{1/\gamma}$. Therefore, for this sequence, the local deviation from the null hypothesis is measured by a distance in terms of the support function, and its magnitude is controlled by the parameter h.

The power of the test has the following properties.

THEOREM 3.10: *Suppose the conditions of Theorem 3.8 hold. Then (i) The test is asymptotically locally unbiased:* $\liminf_{n\to\infty} \pi_{n,b,t}(\theta_n) \geq \alpha$ *for any* $h > 0$*;*

(ii) The limiting power function satisfies

$$\begin{aligned}\liminf_{n\to\infty} \pi_{n,b,t}(\theta_n) &\geq \liminf_{n\to\infty} P_{\theta_n,\Theta_I}\left(\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}_n(\theta,\lambda) > t\right)\\ &\geq P_{\theta_0,\Theta_I}\left(\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}(\theta,\lambda) > t\right),\end{aligned}$$

where

$$\begin{aligned}\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} &= H(p,\Theta_I) \times (K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} \cap L)\\ &= \left\{(\theta,\lambda) : \langle p_0,\theta\rangle = s(p_0,\Theta_I), \langle p_0,\lambda\rangle \geq h - c^{\rightarrow}_{1-\alpha}(p_0,t), \lambda \in L\right\},\end{aligned}$$

for some compact set L*.*

The power depends on the event

$$\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}_n(\theta,\lambda) > t. \tag{3.10}$$

This representation gives several insights. First, as a natural consequence of the construction of the local alternatives, the asymptotic power of the test is fully determined by p_0 and h, given t. Note that the infimum on the left hand side is essentially determined by the set

$$K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} = \{\lambda : \langle p_0, \lambda\rangle \geq h - c^{\rightarrow}_{1-\alpha}(p_0,t)\}.$$

From this, one can see the test has the same power against two distinct alternatives θ_n and θ_n' that are on the same hyperplane with the normal vector p_0, which has the distance $h/a_n^{1/\gamma}$ to θ_0. This is because we reject the hypothesis if and only if the constraint set $K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} \cap L$ is separated from the level-t set of $\tilde{\zeta}_n(\theta_0,\cdot)$. One can show that the probability of the event above is asymptotically α when $h = 0$. As h increases, the constraint set $K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} \cap L$ escapes to the horizon, and the infimum over this set becomes arbitrarily large, which drives $\pi_{n,b,t}(\theta_n)$ to one as $h \to \infty$.

Second, for a given $t \in \mathbb{R}_+$ and $p_0 \in \mathbb{S}^{d-1}$, the slope of the power function as a function of h is determined by the shape of the criterion function. If the econometrician wishes to achieve a rapid increase in power against alternatives that are away from the null with a specific direction p_0, she should choose a criterion function Q_n to force $\tilde{\zeta}_n(\theta_0,\cdot)$ to grow more rapidly in the direction p_0. This makes it more likely for the left hand side of eq. (3.10) to exceed t.

Third, for a given $h \geq 0$ and $t \geq 0$, the power differs with $p_0 \in \mathbb{S}^{d-1}$ depending on the precision with which each support set is estimated. This explains why we use different critical values for different directions when we constructed $\check{\Theta}_{n,b,1-\alpha}(t)$.

An Extension of Pointwise Inference

A simple extension of pointwise inference yields a conservative test for a hypothesis that Θ_I has a nonempty intersection with a known set Θ_0. When Θ_0 is a set of parameter values that satisfy some restrictions, this test can be used to assess the validity of such restrictions. This type of hypothesis has been studied in Romano and Shaikh (2008) for parametric models and Santos (2007) for nonparametric models.

Now let $\Theta_0 \in \mathcal{K}$ be a nonempty closed subset of Θ. Consider testing

$$H_0 : \Theta_0 \cap \Theta_I \neq \varnothing \quad vs. \quad H_1 : \Theta_0 \cap \Theta_I = \varnothing. \tag{3.11}$$

Here, Θ_0 collects parameter values that satisfy the restrictions of interest. The null states that there is at least one element in the identified set satisfying the restrictions. Rejection means that none of the parameters in the identified set satisfy the restrictions, implying that the data generating parameter value does not satisfy the restrictions.

The null hypothesis can be equivalently stated as $H_0 : \exists\theta_0 \in \Theta_0$ such that $\{\theta_0\} \subseteq \Theta_I$. Note that $\inf_{\theta_0\in\Theta_0} \vec{d}_H(\{\theta_0\}, \Theta_I) = 0$ under the null hypothesis, and $\inf_{\theta_0\in\Theta_0} \vec{d}_H(\{\theta_0\}, \Theta_I) > 0$ under the alternative hypothesis. Therefore, a natural test statistic is $\inf_{\theta_0\in\Theta_0} T^{\rightarrow}_{n,\theta_0}(t) = \inf_{\theta_0\in\Theta_0} a_n^{1/\gamma} \vec{d}_H(\{\theta_0\}, \hat{\Theta}_n(t))$. The triangle inequality (3.6) implies

$$\inf_{\theta_0\in\Theta_0} \vec{d}_H(\{\theta_0\}, \hat{\Theta}_n(t)) \leq \inf_{\theta_0\in\Theta_0} \vec{d}_H(\{\theta_0\}, \Theta_I) + \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)).$$

Under the null hypothesis, the first term on the right vanishes. Therefore, we can use the asymptotic conservative critical value $c^{\rightarrow}_{n,b,1-\alpha}(t)$ computed by Algorithm 3.2 to test the hypothesis. As a corollary to Theorem 3.4, results for the rejection probability and the consistency against fixed alternatives follow as before.

COROLLARY 3.5: *Suppose the conditions of Theorem 3.4 hold. Let Θ_0 be a nonempty closed subset of Θ^o.*

(i) If $\Theta_0 \cap \Theta_I \neq \varnothing$ and $F^{\rightarrow}(\cdot, t)$ is continuous and strictly increasing at $c^{\rightarrow}_{1-\alpha}(t)$, then $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t) = c^{\rightarrow}_{1-\alpha}(t) + o_p(1)$ and the test has asymptotic rejection probability bounded above by α :

$$\lim_{n\to\infty} P\left(\inf_{\theta_0\in\Theta_0} T^{\rightarrow}_{n,\theta_0}(t) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right) \leq \alpha;$$

(ii) If $\Theta_0 \cap \Theta_I = \varnothing$, then the test is consistent:

$$\lim_{n\to\infty} P\left(\inf_{\theta_0\in\Theta_0} T^{\rightarrow}_{n,\theta_0}(t) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right) = 1.$$

One may envisage other extensions of pointwise inference. One such possibility is inference on linear functionals of θ_0. This extension is straightforward in our framework, as any linear functional of θ_0 can be represented as $\langle p, \theta_0\rangle$ for some $p \in \mathbb{R}^d$. This may also be extended to nonlinear functionals, but to keep a tight focus here, we leave that analysis to elsewhere.

4 Moment Inequality Models

In this section, we pay special attention to a class of economic models with an identified set defined by finitely many moment inequalities. This class has been extensively studied recently[17]. Leading examples in this class of models are a regression model with censored outcome variables (Manski and Tamer, 2002), entry game models (Ciliberto and Tamer,

[17]Recent research in this area includes Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2005), Rosen (2008), CHT, Fan and Park (2007), Galichon and Henry (2007), BM, Andrews and Guggenberger (2009), Andrews and Soares (2009), Bugni (2009), Canay (2009), Galichon and Henry (2009), Hahn and Ridder (2009), Moon and Schorfheide (2009), and Yıldız (2009).

2009), and dynamic game models (Bajari, Benkard, and Levin, 2007). We first show that this class can be studied within the framework developed above. We provide a set of conditions for this class that ensure the high level assumptions presented in sections 2 and 3. In section 4.2, we provide additional results that can be obtained by using CHT's quadratic criterion function. In particular, we establish the asymptotic equivalence of the squared directed Hausdorff distance statistic and CHT's QLR statistic.

4.1 General Results for Moment Inequality Models

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. We consider functions $m_j : \mathbb{R}^k \times \mathbb{R}^d \to \bar{\mathbb{R}}$, $j = 1, \cdots, J$, that define the following moment inequality restrictions.

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

Interest attaches to the identified set, which comprises the values at which the moment inequalities are satisfied: i.e., $\Theta_I := \{\theta \in \Theta : E\left(m_j(X;\theta)\right) \leq 0, j = 1, \cdots, J\}$.

Let m_θ be a $J \times 1$ vector whose j-th component is $m_{j,\theta} := m_j(X;\theta)$. Let $\mathcal{P}_J$ be the space of symmetric positive definite real-valued $J \times J$ matrices, and let $\bar{\mathcal{P}}_J$ be the space of symmetric positive definite extended real-valued $J \times J$ matrices. For any $\theta \in \mathbb{R}^d$, let $W(\theta) \in \bar{\mathcal{P}}_J$ be a weighting matrix, and let $\{\hat{W}_n : \Omega \times \mathbb{R}^d \to \bar{\mathcal{P}}_J\}$ be a sequence of (possibly random) positive definite weighting matrices. For brevity, we write $\hat{W}_n(\theta)$. We consider population and sample criterion functions of the form:

$$Q(\theta) = \varphi\left(E(m_\theta), W(\theta)\right)$$
$$Q_n(\theta) = \varphi\left(\hat{E}_n(m_\theta), \hat{W}_n(\theta)\right),$$

where $\varphi : \bar{\mathbb{R}}^J \times \bar{\mathcal{P}}_J \to \bar{\mathbb{R}}_+$ is a non-negative continuous function of the moment condition and the weighting matrix. For example, CHT and Romano and Shaikh (2008, 2009) consider the following functional form for Q_n:

$$Q_n(\theta) = \sum_{j=1}^{J} (\hat{W}_{jn}^{1/2}(\theta)\hat{E}_n(m_{j,\theta}))_+^2,$$

where $\hat{W}_{jn}(\theta)$ is the j-th diagonal element of $\hat{W}_n(\theta)$. Manski and Tamer (2002) and Rosen (2008) use the form:

$$Q_n(\theta) = \inf_{\mu \in \mathbb{R}^J_-} (\hat{E}_n(m_\theta) - \mu)'\hat{W}_n(\theta)(\hat{E}_n(m_\theta) - \mu),$$

where $\mathbb{R}^J_- = \{x \in \mathbb{R}^J : x_j \leq 0, j = 1, \cdots, J\}$. We focus on a class of criterion functions that includes the examples above as special cases[18]. We assume the following regularity conditions on the parameter space, the moment conditions, and the "index function" φ.

Assumption 4.1: *Let $J \in \mathbb{N}$. $\varphi : \bar{\mathbb{R}}^J \times \bar{\mathcal{P}}_J \to \bar{\mathbb{R}}_+$ is a non-negative continuous function such that for any $w \in \mathcal{P}_J$, $\varphi(y, w) = 0$ if and only if $y \leq 0$, i.e. $y_j \leq 0$ for $j = 1, \cdots, J$, and $\varphi(y, w) = \infty$ if y or w contains an infinite element. Let $\Theta \subset \mathbb{R}^d, d \in \mathbb{N}$, be compact and convex with nonempty interior. Let $W : \mathbb{R}^d \to \bar{\mathcal{P}}_J$ be a measurable mapping, and suppose that W is finite and continous on Θ and that if $\theta \notin \Theta$ then $\det(W(\theta)) = \infty$. Let $k \in \mathbb{N}$; for each $j = 1, \cdots, J$, $m_j : \mathbb{R}^k \times \mathbb{R}^d \to \bar{\mathbb{R}}$ is jointly measurable, and for each $x \in \mathbb{R}^k$, if $\theta \notin \Theta$ then $m_j(x, \theta) = \infty$. Let $(\Omega, \mathfrak{F}, P)$ be a complete probability space. Let $\{X_i : \Omega \to \mathbb{R}^k\}$ be a sequence of identically distributed random vectors such that for each $\theta \in \Theta$ and $j = 1, \cdots, J$, $E(m_j(X_i, \theta)) < \infty$. Let $\hat{W}_n : \Omega \times \mathbb{R}^d \to \bar{\mathcal{P}}_J$ be jointly measurable, and suppose that for each $\omega \in \Omega$, $\hat{W}_n(\omega, \cdot)$ is finite and continuous on Θ, uniformly in n, and for each $\omega \in \Omega$, if $\theta \notin \Theta$ then $\det \hat{W}_n(\omega, \theta) = \infty$. Define $Q(\theta) := \varphi\left(E(m_\theta), W(\theta)\right)$ and $Q_n(\theta) := \varphi(\hat{E}_n(m_\theta), \hat{W}_n(\theta))$.*

Assumption 4.1 ensures that Assumption 2.1 holds for moment inequality models. The assumed continuity of W on Θ and its behavior outside of Θ ensures that its minimum eigenvalue is bounded from below by a positive constant over $\mathbb{R}^d$. The almost sure properness of the sample criterion function Q_n is ensured by the requirements that φ is a nonnegative function and that $E(m_{j,\theta})$ and $W(\theta)$ are finite on Θ. Using the criterion function Q, the identified set can be defined as $\Theta_I = \{\theta : Q(\theta) = 0\}$.

The following condition ensures Assumption 2.2.

Assumption 4.2: *(i) There exists $\theta \in \Theta$ such that $E(m_{j,\theta}) \leq 0$ for $j = 1, \cdots, J$. The map $\theta \mapsto \varphi(E(m_\theta), W(\theta))$ is continuous and convex on Θ; (ii) $\{\theta \in \Theta : \varphi(E(m_\theta), W(\theta)) = 0\} \subset \Theta^o$.*

Assumption 4.2 (i) ensures nonemptiness, closedness, and convexity of the identified set. Assumption 4.2 (ii) ensures that the identified set is in the interior of Θ.

Conditions required for the consistency of the set estimator $\hat{\Theta}_n(t)$ are standard.[19] In particular, we must ensure the uniform convergence of Q_n. The rate of convergence depends on the choice of the index function φ. Here, we give primitive conditions on the moment conditions and the index function based on CHT's condition M.2. For this, we introduce the ϵ-contraction of Θ_I, which is defined by $\Theta_I^{-\epsilon} := \{\theta \in \Theta_I : d(\theta, \Theta \setminus \Theta_I) \geq \epsilon\}$ for $\epsilon > 0$.

Assumption 4.3: *(i-a) There exist $0 < L_1 < \infty$ and a continuous increasing function*

[18] It would be interesting to extend our analysis here to a more general class within which we can also study moment equality models. Such a general class was considered in Andrews and Guggenberger (2009) and Andrews and Soares (2009).

[19] Strictly speaking, one needs to establish the measurability of $d_H(\hat{\Theta}_n(t), \Theta_I)$ to discuss consistency. It is known that the measurability of $\hat{\Theta}_n(t)$ as a random closed set is sufficient for this purpose. For details about the measurability of level set estimators, see Kaido and White (2008).

$h_1 : \bar{\mathbb{R}}_+ \to \bar{\mathbb{R}}_+$ such that for any $w \in \mathcal{P}_J$ and $x, x^* \in \mathbb{R}^J$, $|\varphi(x,w) - \varphi(x^*,w)| \le L_1 h_1(\|x - x^*\|)$, and there exist $0 < L_2 < \infty$ and a continuous increasing function $h_2 : \bar{\mathbb{R}}_+ \to \bar{\mathbb{R}}_+$ such that for any $x \in \mathbb{R}^J$ and $w, w^* \in \mathcal{P}_J$, $|\varphi(x,w) - \varphi(x,w^*)| \le L_2 h_2(\max_{i,j} |w_{ij} - w^*_{ij}|)$; (i-b) $\{m_\theta : \theta \in \Theta\}$ is a P-Donsker class, and $\hat{W}_n(\theta) - W(\theta) = o_p(1)$ uniformly over Θ; (ii-a) $\sup_{\Theta_I} Q_n(\theta) = O_p(1/a_n)$; (ii-b) There exist positive constants (C_1, δ) such that for any $\theta \in \Theta$, $\|E(m_\theta)\|_+ \ge C_1 \min\{d(\theta, \Theta_I), \delta\}$; (ii-c) There exist positive constants (C_2, γ) such that for any $\omega \in \mathcal{P}_J$ and $x \in \mathbb{R}^J$, $\varphi(x,w) \ge C_2 \|x\|_+^\gamma$; (ii-d) There exist positive constants $(C_3, C_4, \bar{\epsilon})$ such that for any $0 \le \epsilon \le \bar{\epsilon}$ and $\theta \in \Theta_I^{-\epsilon}$, $\max_{1 \le j \le J} E(m_{j,\theta}) \le -C_3 \epsilon$, and $d_H(\Theta_I^{-\epsilon}, \Theta_I) \le C_4 \epsilon$.

Assumptions 4.3 (i-a,b) are sufficient for the uniform convergence of Q_n on Θ[20]. Assumption 4.3 (ii) collects conditions necessary for the convergence rate result. Condition (ii-a) requires the sample criterion function Q_n to vanish over the identified set at a rate of $1/a_n$. Assumption 4.3 (ii-b) requires the norm of $E(m_\theta)$ to be bounded from below by the distance from the identified set when θ is outside Θ_I. Together with Assumption 4.3 (ii-c), this ensures the existence of a polynomial minorant, which is required in Assumption 2.3 (ii). Assumption 4.3 (ii-d) requires the moment conditions to take strictly negative values on the contracted identified set. This enables us to approximate the identified set by its contraction $\Theta_I^{-\epsilon}$, on which the sample criterion function $a_n Q_n(\theta)$ vanishes. As CHT illustrate, Assumption 4.3 (ii-d) holds in many applications. This condition implies Assumption 2.4, which suffices to attain the exact rate of convergence $a_n^{1/\gamma}$ without setting $t \ge \sup_\Theta a_n Q_n(\theta)$.

The next step is to show that $\tilde{\zeta}_n(\theta, \lambda) = n Q_n(\theta + \lambda / a_n^{1/\gamma})$ satisfies the local process regularity conditions given in Assumption 3.1. Most importantly, we need to ensure that $\tilde{\zeta}_n$ weakly epiconverges to a well-defined limit. To illustrate the key ideas, we take a slightly generalized version of CHT's criterion function as an example.

Let $x \circ y$ denote the entrywise (Hadamard) product of $x, y \in \mathbb{R}^J$. Let $s : \mathbb{R}^J \to \{1,0\}^J$ be a vector-valued mapping whose j-th component is $s_j = 1\{x_j > 0\}$. Let the index function be defined by $\varphi(x,w) := \left\|w^{1/2}x\right\|_+^2 := \|w^{1/2}(x \circ s)\|^2$. The sample criterion function is then $Q_n(\theta) = \|\hat{W}_n^{1/2}(\theta)\hat{E}_n(m_\theta)\|_+^2$. As the weighting matrix need not be diagonal, this is a slightly generalized version of the criterion function used by CHT.

With this choice of index function, we can take $a_n = n$ and $\gamma = 2$. That is, $nQ_n(\theta)$ has nondegenerate asymptotics, and $\sqrt{n} d_H(\hat{\Theta}_n(t), \Theta_I) = O_p(1)$, given Assumptions 4.1, 4.2, and 4.3. Suppose that m_θ allows a first-order expansion $m_{\theta^*} = m_\theta + \nabla' m_\theta (\theta^* - \theta) + o(|\theta^* - \theta|)$ on Θ^o, where ∇m_θ is a d-by-J matrix and $o(|\theta^* - \theta|)$ represents a small order term. Under

[20]If we further assume that $\theta \mapsto \varphi(\hat{E}_n m_\theta, \hat{W}_n(\theta))$ is globally convex on Θ, a weaker assumption that $\hat{E}_n(\theta)$ and $W_n(\theta)$ converge in probability pointwise is sufficient as in Andersen and Gill (1982) Corollary II.2 and Newey and McFadden (1992) Theorem 2.7.

these assumptions, we can write

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

where we define $\mathbb{G}_n := \sqrt{n}(\hat{E}_n - E)$ and $\mathcal{M}_n(\theta,\lambda) = \mathbb{G}_n m_\theta + \hat{E}_n(\nabla' m_\theta)\lambda + \sqrt{n}E(m_\theta)$.

By the P-Donsker property of the moment functions, $\mathbb{G}_n m_\theta \overset{u.d.}{\to} \mathbb{G}(\theta)$ in $l^\infty(\Theta)$, where $\mathbb{G}$ is a $J\times 1$ zero-mean Gaussian process with almost surely continuous paths, and $Var(\mathbb{G}_j(\theta)) > 0$ for each $\theta\in\Theta$ and $j = 1,\cdots,J$. Together with the P-Donsker property, a set of general assumptions is often available to ensure that, for each $(\theta,\lambda)\in\Theta^o\times\mathbb{R}^d$, $\mathcal{M}_n(\theta,\lambda) \overset{f.d.}{\to} \mathcal{M}(\theta,\lambda) := \mathbb{G}(\theta)+\Pi(\theta)\lambda+\varsigma(\theta)$ and $\hat{W}_n(\theta+\lambda/\sqrt{n}) \overset{p}{\to} W(\theta)$, where $\varsigma(\theta)$'s j-th component satisfies

$$\varsigma_j(\theta) = \begin{cases} -\infty & \text{if } E(m_{j,\theta}) < 0 \\ 0 & \text{if } E(m_{j,\theta}) = 0 \\ \infty & \text{if } E(m_{j,\theta}) > 0 \end{cases} \tag{4.1}$$

for $j = 1,2,\cdots,J$. The components of $\varsigma(\theta)$ are unbounded if the corresponding population moment inequalities are not binding, but the truncation operator $(\cdot)_+$ makes the criterion function always bounded from below by 0, which ensures the properness of the limiting process.

Similarly, for general choice of φ, one can often show $\tilde{\zeta}_n(\theta,\lambda) = \varphi(\mathcal{M}_n(\theta,\lambda),\hat{W}_n(\theta+\lambda/a_n^{1/\gamma})) + o_p(1)$ for each θ and λ. Then the continuous mapping theorem implies

$$\tilde{\zeta}_n(\theta,\lambda) \overset{f.d.}{\to} \varphi\left(\mathcal{M}(\theta,\lambda), W(\theta)\right)$$

Recall that, to apply Theorem 3.1, we need to establish the weak epiconvergence of $\tilde{\zeta}_n$ instead of the weak finite-dimensional convergence. Provided that the weak finite-dimensional limit exists, Knight (1999) shows that the weak finite-dimensional limit is also the weak epi-limit if and only if the sequence $\tilde{\zeta}_n(\theta,\lambda)$ is equi-lower-semicontinuous[21]. A general sufficient condition that ensures the desired weak epiconvergence and other local process regularities is the following.

Assumption 4.4: *(i-a) For each $j = 1,\cdots,J$, and $x\in\mathbb{R}^k$, $m_j(x,\cdot)$ is continuously differentiable with respect to θ on Θ^o with a continuous gradient $\nabla m_\theta(x,\cdot)\in\mathbb{R}^{d\times J}$, and*

[21]The mathematical appendix summarizes Knight (1999)'s results. When $\tilde{\zeta}_n(\theta,\lambda)$ is also globally convex, it suffices to check that the limiting function is finite on some open set. See Geyer (2003) for details.

for some continuous mapping $\Pi : \Theta^o \mapsto \mathbb{R}^{J\times d}$ *and each* θ *in* Θ^o, $\hat{E}_n(\nabla' m_\theta) = \Pi(\theta) + o_p(1)$; *(i-b)* $\sqrt{n}E(m_\theta) = \varsigma(\theta) + o_p(1)$ *for each* θ *in* Θ, *where* ς *is defined by Eq. (4.1); (ii) The map* $\theta \mapsto \varphi(\hat{E}_n(m_\theta), \hat{W}_n(\theta))$ *is convex in a neighborhood of* Θ_I*; (iii) The map* $(\theta, \lambda) \mapsto \varphi(\mathcal{M}_n(\theta,\lambda), \hat{W}_n(\theta + \lambda/a_n^{1/\gamma}))$ *is equi-lower-semicontinuous on* $\Theta^o \times \mathbb{R}^d$.

The conditions in Assumption 4.4 are plausibly general. In addition to the P-Donskerness of $\{m_\theta : \theta \in \Theta\}$, we only require the finite-dimensional pointwise convergence of other terms in $\mathcal{M}_n(\theta,\lambda)$. A standard LLN will ensure this requirement.

Next, the following theorem establishes Assumptions 2.1-2.4, and the local process regularity (Assumption 3.1), including weak epiconvergence.

Theorem 4.1: *Suppose Assumptions 4.1, 4.2 4.3, and 4.4 hold.*

Then Assumptions 2.1, 2.2, 2.3, 2.4 and 3.1 are satisfied with weak epilimit $\tilde{\zeta}(\theta,\lambda) := \varphi\left(\mathcal{M}(\theta,\lambda), W(\theta)\right)$.

Theorem 3.1 now applies. An important corollary is the following.

Corollary 4.1: *Suppose Assumptions 4.1, 4.2 4.3, and 4.4 hold.*

Then $\sqrt{n}d_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \|\mathcal{Z}(\cdot,t)\|_{\mathcal{C}(\mathbb{S}^{d-1})}$, *and* $\sqrt{n}\vec{d}_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(\cdot,t)\}_+$, *where* $\mathcal{Z}(\cdot,t)$ *can be represented as*

$$\mathcal{Z}(p,t) = \sup_{\theta\in H(p,\Theta_I)} \sup_{\lambda\in\{\lambda:\varphi(\mathcal{M}(\theta,\lambda),W(\theta))\leq t\}} \langle p, \lambda\rangle. \tag{4.2}$$

The representation above specifies how the limiting process $\mathcal{Z}(\cdot,t)$ depends on the weak epilimit $\varphi(\mathcal{M}(\theta,\lambda), W(\theta))$. Note that the asymptotic distribution of $\mathcal{Z}(\cdot,t)$ depends non-trivially on the identified set Θ_I.

4.2 A Closed Form for the Limiting Process and the Equivalence of Wald and QLR Statistics

In the previous section, we provided general conditions for moment inequality models that ensure the high level assumptions in section 3. In this section, we develop further results that rely on the properties of CHT's quadratic criterion function.

The goal of this section is to show that (i) a closed form for the limiting process $\mathcal{Z}(\cdot,t)$ can be derived; (ii) for each p, the limiting process $\mathcal{Z}(p,t)$ depends only on the active moment inequalities at $\theta \in H(p,\Theta_I)$; (iii) a certain choice of weighting matrix $W(\theta)$ makes the limiting process take the form $\mathcal{Z}(p,t) = \mu(t) + \mathcal{Z}^*(p)$; and (iv) The Wald statistic (squared directed Hausdorff distance) and CHT's QLR statistic are asymptotically equivalent, under this choice of the weighting matrix and some additional assumptions.

We introduce some further notation to denote active and slack moment inequalities. For each $\theta \in \partial\Theta_I$, let $\mathcal{J}(\theta) \subseteq \{1, \cdots, J\}$ be the set of indices associated with active moment inequalities, i.e., $E(m_{j,\theta}) = 0$ for all $j \in \mathcal{J}(\theta)$. We denote by $J(\theta)$ the number of elements in $\mathcal{J}(\theta)$. Similarly, let $\mathcal{J}^c(\theta) \subseteq \{1, \cdots, J\}$ collect indices associated with slack moment inequalities at $\theta \in \partial\Theta_I$, i.e., $E(m_{j,\theta}) < 0$ for all $\mathcal{J}^c(\theta)$.

Let $\Pi_{\mathcal{J}(\theta)}(\theta)$ denote the $J(\theta) \times d$ matrix that stacks rows of $\Pi(\theta)$ whose indices belong to $\mathcal{J}(\theta)$. Similarly, 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 $\mathcal{J}(\theta)$. Let $W_{\mathcal{J}(\theta)}$ denote the $J(\theta) \times J(\theta)$ matrix that collects (i, j) elements of $W(\theta)$ for $i, j \in \mathcal{J}(\theta)$.

We consider the following problem, which is a part of the optimization problem that defines $\mathcal{Z}(\cdot, t)$ in Eq. (4.2), while fixing $p \in \mathbb{S}^{d-1}$, $\theta \in H(p, \Theta_I)$, and $t \in \mathbb{R}_+$.

$$\begin{aligned} \sup_{\lambda} \quad & \langle p, \lambda \rangle \\ s.t. \quad & \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)[\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda]\|^2_+ \le t. \end{aligned} \tag{4.3}$$

Note that the constraint involves only selected rows of $\mathcal{M}(\theta, \lambda)$ whose indices are in $\mathcal{J}(\theta)$. This is because $\varsigma_j(\theta) = -\infty$ if $E(m_{j,\theta}) < 0$, and the index function φ truncates such components. The rows of $\mathcal{M}(\theta, \lambda)$ with indices belonging to $\mathcal{J}^c(\theta)$, therefore, do not marginally affect the constraint. Note also that the $\varsigma_j(\theta)$'s no longer appear in the constraint because $\varsigma_j(\theta) = 0$ for $j \in \mathcal{J}(\theta)$.

To obtain a closed form for $\mathcal{Z}(\cdot, t)$, we assume the following further conditions.

Assumption 4.5: *(i) 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; (ii) For each $\theta \in \partial\Theta_I$ and $p \in \mathbb{S}^{d-1}$, there exists a vector $\eta \in \mathbb{R}^{J(\theta)}_+ \setminus \{0\}$ such that $p = \Pi'_{\mathcal{J}(\theta)}\eta$.*

Assumption 4.5 (i) is a linear independence constraint qualification condition. This ensures the solution to the problem in eq (4.3) satisfies the Karush-Kuhn-Tucker (KKT) conditions given in the mathematical appendix. Assumption 4.5 (ii) is not restrictive, as it usually holds as a necessary condition for the following auxiliary optimization problem, which can be used to characterize the boundary points of the identified set:

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

Using these additional assumptions, we can explicitly solve the optimization problem in Eq. (4.3) to obtain the following result.

Corollary 4.2: *Suppose the conditions of Theorem 4.1 and Assumption 4.5 hold. Suppose $\varphi(x, w) = \|w^{1/2}x\|^2_+$. Then the process $\mathcal{Z}(\cdot, t)$ in Corollary 4.1 can be represented as*

$$\mathcal{Z}(p, t) = \sup_{\theta \in H(p, \Theta_I)} \left\{ \|\mathcal{R}(p, \theta)\|\, t^{1/2} - \left\langle \mathcal{R}(p, \theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)} \right\rangle \right\}, \tag{4.4}$$

where

$$\mathcal{R}(p,\theta) := W_{\mathcal{J}(\theta)}^{-1/2}\left(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)'\right)^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p.$$

Furthermore, suppose $W(\theta)$ *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 the limiting process takes the form* $\mathcal{Z}(p,t) = \mu(t) + \mathcal{Z}^*(p)$ *with* $\mu(t) = t^{1/2}$ *and* $\mathcal{Z}^*(p) = \sup_{\theta \in H(p,\Theta_I)} -\langle [\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)']^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p, \mathbb{G}_{\mathcal{J}(\theta)}(\theta)\rangle$.

Equation 4.4 shows the limiting process $\mathcal{Z}(\cdot,t)$ depends on the multivariate Gaussian process $\mathbb{G}$, but again we note that the only selected components of $\mathbb{G}$ are relevant. Therefore, for each $p \in \mathbb{S}^{d-1}$, the asymptotic distribution of the normalized support function depends only on the active moment inequalities at each boundary point of the identified set. This is a common feature of the statistics studied in the literature (e.g. Rosen, 2008, Andrews and Soares, 2009).

If the weighting matrix satisfies $W_{\mathcal{J}(\theta)}(\theta) := [\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)']^{-1}$ at each boundary point, then straightforward algebra shows $\|\mathcal{R}(p,\theta)\| = 1$, which makes the first term in Eq. (4.4) independent of θ. With this choice of weighting matrix, the limiting process takes the form $\mathcal{Z}(p,t) := t^{1/2} + \mathcal{Z}^*(p)$[22].

We now make use of the representation result above to compare the weak limit of the Wald statistic with that of CHT's QLR statistic: $\sup_{\Theta_I} a_nQ_n(\theta)$. The QLR statistic can be written as

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

As the second term on the right hand side asymptotically vanishes by Assumption 4.3 (ii-d), it suffices to study the first term. Using the local process $\tilde{\zeta}_n$, define

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

where $K^-_{u,p} := \{\lambda \in \mathbb{R}^d : \langle p,\lambda\rangle \le u\}$. Note that $\sup_{p\in\mathbb{S}^{d-1}} \mathcal{L}_n(p,0) = \sup_{\theta\in\partial\Theta_I} a_nQ_n(\theta)$. We therefore study the asymptotic behavior of the process $\mathcal{L}_n(\cdot,u)$ to study that of the QLR statistic. The following theorem establishes the weak convergence of $\mathcal{L}_n(\cdot,u)$. The regularity conditions for this theorem are given in the mathematical appendix.

Theorem 4.2: *Suppose the conditions of Corollary 4.2 hold. Suppose Assumption A.1*

[22] In sample, one may use a sample analog $\hat{W}_{n,\mathcal{J}_n(\theta)}(\theta) := (\hat{E}_{n,\mathcal{J}_n(\theta)}[\nabla m_\theta]\hat{E}_{n,\mathcal{J}_n(\theta)}[\nabla m_\theta]')^{-1}$ to construct Q_n. Here, for each n, $\mathcal{J}_n(\theta)$ is a mapping from Θ to a subset of $\{1,\cdots,J\}$ that selects (approximately) binding sample moment conditions at θ. Such moment selection mechanisms are studied in Andrews and Soares (2009).

holds. Then $\mathcal{L}_n(\cdot, u) \stackrel{u.d.}{\to} \mathcal{L}(\cdot, u)$ *for each* u*, and the process* $\mathcal{L}$ *can be represented as*

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

Based on this theorem, the following corollary establishes two equivalence results. The first result is the equivalence of the distributional limits of the Wald and the QLR statistics. The second result is the equality of the levels of the criterion function used by the Wald approach and the QLR approach to construct confidence sets. Recall that $t^*_{1-\alpha} := \inf\{t : P(\sup_{p \in \mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\} \leq 0) \geq 1-\alpha\}$, and $\tau^*_{1-\alpha}$ is the asymptotic $1-\alpha$ quantile of the QLR statistic.

Corollary 4.3 (Asymptotic Equivalence for Moment Inequalities): *Suppose the conditions of Theorem 4.2 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 the boundary point of* Θ_I *such that* $H(p,\Theta_I) = \{\theta_I(p)\}$*.*

Then, (i)

$$\sup_{p \in \mathbb{S}^{d-1}} \{-\mathcal{Z}_n(p,t) + t^{1/2}\}^2_+ \stackrel{d}{\to} \mathbf{Z} \quad \textit{and} \quad \sup_{\Theta_I} nQ_n(\theta) \stackrel{d}{\to} \mathbf{Z},$$

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) $t^*_{1-\alpha} = \tau^*_{1-\alpha}$*.*

Corollary 4.3 shows that our 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, the supremum of a truncated and squared Gaussian process. The second result also has important conseqeuces. It implies the asymptotic equivalence of the Wald and QLR confidence sets for Θ_I. This is due to Theorem 3.6. When $t^*_{1-\alpha} = \tau^*_{1-\alpha}$, Theorem 3.6 implies

$$d_H\left(\hat{\Theta}_n^{\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)}(t), \hat{\Theta}_n(\tau^*_{1-\alpha})\right) = o_p(a_n^{-1/\gamma}),$$

for any $0 \leq t \leq \tau^*_{1-\alpha}$. The first argument of d_H on the left hand side is the Wald confidence set, which is an expansion of the set estimator. The second argument is the QLR confidence set, which is a level set that uses an asymptotic quantile of the QLR statistic as a level. Despite the fundamental difference in ways these confidence sets are constructed, they are asymptotically equivalent in terms of the Hausdorff metric. These are fundamental results that establish the relationship between the Wald and QLR approaches.

5 Examples, Monte Carlo Experiments, and Applications

5.1 Examples

In this section, we analyze two examples studied in the literature using our inference method. The first model has an identified set that is a closed interval. Using this example, we illustrate our equivalence results in more detail and give a new interpretation to the results established by BM.

Example 5.1 (Interval Identified Model): *Let X be an unobserved random variable with mean $\theta = E(X)$. Let X_1 and X_2 be observable random variables that satisfy the moment inequalities $E(X_1) \leq \theta \leq E(X_2)$.*

Let $\theta_1 = E(X_{1i})$ and $\theta_2 = E(X_{2i})$. The identified set for θ is a closed interval $\Theta_I = [\theta_1, \theta_2]$. Following the analysis in section 4, Θ_I can be characterized as a set of minimizers of the criterion function

$$Q(\theta) = \|W(\theta)^{1/2} E(m_\theta)\|_+^2,$$

where $m_\theta = (X_1 - \theta, \theta - X_2)'$. Define the sample criterion function by

$$Q_n(\theta) = \|\hat{W}_n(\theta)^{1/2} \hat{E}_n(m_\theta)\|_+^2.$$

For simplicity, we set $W(\theta)$ and $\hat{W}_n(\theta)$ to the identity matrix.

It is straightforward to show that these population and sample criterion functions satisfy Assumptions 4.1, 4.2, 4.3, and 4.4. The following results follow immediately from Corollaries 4.1 and 4.2,

Corollary 5.1: *Let ι be a 2-by-1 vector of ones. Let $t \in \mathbb{R}_+$. Suppose*

$$(\sqrt{n}(\hat{E}_n(X_{1i}) - \theta_1), \sqrt{n}(\hat{E}_n(X_{2i}) - \theta_2))' \xrightarrow{d} N(0, \Omega)$$

and $-\infty < \theta_1 < \theta_2 < \infty$.

Then $\sqrt{n} d_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \max\{|\mathcal{Z}(-1,t)|, |\mathcal{Z}(1,t)|\}$ and $\sqrt{n}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \xrightarrow{d} \max\{-\mathcal{Z}(-1,t)_+, -\mathcal{Z}(1,t)_+\}$, where $\mathcal{Z}(p,t)$ is a Gaussian process on $\mathbb{S}^0 = \{-1, 1\}$ with mean $t^{1/2}\iota$ and covariance kernel $E[\mathcal{Z}(-1,t)\mathcal{Z}(-1,t)] = \Omega_{11}$, $E[\mathcal{Z}(1,t)\mathcal{Z}(1,t)] = \Omega_{22}$, and $E[\mathcal{Z}(-1,t)\ \mathcal{Z}(1,t)] = -\Omega_{12}$.

This result is closely related to that presented by BM (Theorem 3.1), which shows that the normalized support function of their set average estimator weakly converges to a zero-mean Gaussian process that has the same covariance kernel as $\mathcal{Z}(\cdot, t)$. In fact, if we set $t = 0$, the level set estimator is analytically identical to their set-average estimator for this class of problems. An additional interesting result is that, under this choice of the weighting matrix, the squared directed Hausdorff distance is asymptotically equivalent to CHT's QLR-statistic.

We summarize these equivalence results as follows:

THEOREM 5.1: *Let the assumptions of Theorem 5.1 hold. Let* $\mathcal{W}_n := \sqrt{n}\vec{d}_H(\Theta_I, \hat{\Theta}_n(0))$. *Let* $\mathcal{QLR}_n := \sup_{\theta\in\Theta_I} nQ_n(\theta)$ *be CHT's QLR statistic. Let* $\tilde{\mathcal{W}}_n := \sqrt{n}\vec{d}_H(\Theta_I, \tilde{\Theta}_n)$ *be BM's Wald statistic, where* $\tilde{\Theta}_n = n^{-1}\bigoplus_{i=1}^n F_i$ *and* $F_i = [X_{1i}, X_{2i}]$ *for* $i = 1, \cdots, n$. *Let* $\mathcal{Z}$ *be the process given in Corollary 5.1. Then*

$$\mathcal{W}_n^2 \xrightarrow{d} \max\{(-\mathcal{Z}(-1,0))_+^2, (-\mathcal{Z}(1,0))_+^2\} \tag{5.1}$$

$$\mathcal{QLR}_n \xrightarrow{d} \max\{(-\mathcal{Z}(-1,0))_+^2, (-\mathcal{Z}(1,0))_+^2\} \tag{5.2}$$

$$\tilde{\mathcal{W}}_n^2 \xrightarrow{d} \max\{(-\mathcal{Z}(-1,0))_+^2, (-\mathcal{Z}(1,0))_+^2\}. \tag{5.3}$$

The asymptotic equivalence of CHT's QLR statistic and BM's Wald statistic in equations (5.2) and (5.3) is due to BM's Theorem 3.1. Here, Theorem 5.1 adds eq. (5.1).

As we have seen in the previous section, the squared directed Hausdorff distance becomes asymptotically equivalent to CHT's QLR statistic when the weighting matrix satisfies the conditions of Corollary 4.3. As the identity matrix satisfies these, the asymptotic equivalence of $\mathcal{W}_n^2$ and $\mathcal{QLR}_n$ follows[23]. Further, for this example, the set-average estimator is a set of minimizers of the truncated squared loss function; this therefore becomes a level-set estimator with $t = 0$. Thus, the "exact" equivalence of $\mathcal{W}_n^2$ and $\tilde{\mathcal{W}}_n^2$ holds. In sum, the asymptotic equivalence result formerly presented by BM can be understood as a combination of (i) the asymptotic equivalence of the Wald statistic and the QLR statistic for the class of moment inequality models and (ii) the equivalence of the level-set estimator and the set-average estimator under the specific choice of criterion function.

In this example, we may interpret BM's set-average estimator as a set-valued quasi maximum likelihood estimator (QMLE) of Θ_I, where the quasi-log likelihood function is the truncated squared loss used by CHT. This is analogous to the point identified case, where the sample average is the QMLE for the location parameter, under the specification that the data are randomly sampled from a normal distribution, which gives a squared error loss function. It is of interest to extend this notion to a more general class of problems.

The second example studies a regression model with interval-valued outcome variables. Our Monte Carlo experiments will be based on this example.

EXAMPLE 5.2 (Regression with Interval-Censored Outcome): *Let* $\theta \in \Theta \subset \mathbb{R}^d$. *Consider the DGP*:

$$Y_i = X_i'\theta + \epsilon_i, \quad i = 1, 2, \cdots, n,$$

where $E[\epsilon_i|X_i] = 0$ *for all* $i = 1, \cdots, n$. *The outcome variable* Y_i *is not observed but the*

[23] Here, we use $W(\theta) = I_2$, the identity matrix. Since the two constraints don't bind at the same time, the weighting matrix for the equivalence should satisfy $W_1(\theta_1) = (\Pi_1(\theta_1)\Pi_1(\theta_1)')^{-1} = 1$ and $W_2(\theta_1) = (\Pi_2(\theta_2)\Pi_2(\theta_2)')^{-1} = 1$. Obviously, the identity matrix satisfies this condition.

outcome interval $[Y_{1i}, Y_{2i}]$ *is observed for each* $i = 1, \cdots, n$. *The outcome interval satisfies the following moment inequalities*

$$E[Y_{1i}|X_i] \le X_i'\theta \le E[Y_{2i}|X_i], \quad a.s.$$

The identified set can be defined as the set of minimizers of the following criterion function:

$$Q(\theta) := \int (E(Y_{1i}|X_i = x) - x'\theta)_+^2 + (x'\theta - E(Y_{2i}|X_i = x))_+^2 dP(x).$$

Following Chernozhukov, Hong, and Tamer (2004), we use the following minimum-distance type sample criterion function

$$Q_n(\theta) := \frac{1}{n}\sum_{i=1}^{n}(\hat{E}_n(Y_1|X_i) - X_i'\theta)_+^2 + (X_i'\theta - \hat{E}_n(Y_2|X_i))_+^2,$$

where $\hat{E}_n(Y_1|X_i)$ and $\hat{E}_n(Y_2|X_i)$ are estimators of $E(Y_{1i}|X_i)$ and $E(Y_{2i}|X_i)$, respectively.

In our Monte Carlo experiments, we study the cases where X_i is a vector of discrete random variables supported on $\{x_1, \cdots, x_J\}, J \in \mathbb{N}$. For these cases, we use $\hat{E}_n(Y_1|X_i = x_j) = n_j^{-1}\sum_{i:X_i=x_j} Y_{1i}$ and $\hat{E}_n(Y_2|X_i = x_j) = n_j^{-1}\sum_{i:X_i=x_j} Y_{2i}$ as estimators, where $n_j = \sum_i 1\{X_i = x_j\}$.

The sample criterion function can be alternatively written as

$$Q_n(\theta) = \|\hat{W}_n^{1/2}(\theta)\hat{E}_n(m_\theta)\|_+^2,$$

where m_θ is a $2J$-dimensional vector whose components are

$$m_{j,\theta} := \begin{cases} \frac{n}{n_j}(Y_{1i} - X_i'\theta)1_{\{X_i=x_j\}} & \text{for } j = 1, \cdots, J \\ \frac{n}{n_j}(X_i'\theta - Y_{2i})1_{\{X_i=x_j\}} & \text{for } j = J+1, \cdots, 2J. \end{cases}$$

The weighting matrix $\hat{W}_n(\theta)$ is a $2J \times 2J$ diagonal matrix whose j-th diagonal element is n_j/n. Therefore, the example above can be studied within the framework presented in section 4. It can be shown that this model also satisfies the conditions of Theorem 4.1. We use this example to evaluate our inference method.

5.2 Monte Carlo Experiments

We conduct a Monte Carlo experiment using Example 5.2 to examine the performance of our inference methods. In both designs, the regressor $X_i = (1, X_{2i})$ consists of a constant and a real random variable X_{2i}.

We use the same data as Chernozhukov, Hong, and Tamer (2004). The original data is taken from March 2000 wave of the Current Population Survey (CPS) with 13,290 observations on income and education[24]. Chernozhukov, Hong, and Tamer (2004) bracketed each individual's log-income into 15 different categories. We let Y_{1i} (Y_{2i}) be the lower (upper) bound of the bracketed log-income of each individual. Chernozhukov, Hong, and Tamer (2004, sec 4) provide further details of the construction of the data. We randomly draw samples of size $n = 1,000$ or $2,000$ from this CPS population and check the coverage probabilities of the Wald confidence set under different values of the subsample size b.

For this exercise, we use a grid of points $\{p_l = (\cos(w_l), \sin(w_l))', l = 1, \cdots, L\}$ with $L = 100$, where each w_l is taken from an equally spaced grid of points on the interval $[0, 2\pi]$. We use Algorithm 2.1 to compute the support function of the set estimator. This algorithm is quite fast. For samples of size $n = 1,000$ and $2,000$, it takes only 0.051 and 0.054 seconds respectively to compute the support function and approximate boundary $\{\hat{\theta}_n(p_l, t), l = 1, \cdots, L\}$ of the set estimator $\hat{\Theta}_n(t)$[25].

The initial choice of the level t is made in a similar manner to Chernozhukov, Hong, and Tamer (2004). First, we consider an auxiliary point-identified model, where the lower and upper bounds for the individual log-income are $\tilde{Y}_{1i} = \tilde{Y}_{2i} = (Y_{1i} + Y_{2i})/2$ for any i. We use the criterion function Q_n applied to the data $\{(\tilde{Y}_{1i}, \tilde{Y}_{2i}), i = 1, \cdots, n\}$ and compute quantiles of the statistic $n(Q_n(\theta_0^a) - Q_n(\hat{\theta}_n^a))$, where θ_0^a is the minimizer of the population criterion function and $\hat{\theta}_n^a$ its estimator. Let $\hat{t}_{a_0}$ be the $100\times a_0\%$ quantile of the statistic. Chernozhukov, Hong, and Tamer (2004) recommends using t_{a_0} with properly chosen a_0 as an initial level. If $\hat{t}_{a_0}$ is too large, the resulting set estimator may not be expanded because its coverage probability is likely to exceed $1 - \alpha$. Therefore, it is desirable to set a_0 to a value less than the nominal level $1 - \alpha$.

Table 1 reports the coverage probabilities of the Wald confidence set under different values of b and a_0. We also include the coverage probabilities of the QLR confidence set reported by Chernozhukov, Hong, and Tamer (2004) for comparison. The nominal level is $1 - \alpha = 0.95$. The QLR benchmark is reported in the third row. We first set $a_0 = 0.5$ (median) for choosing the initial level. Overall, the Wald confidence set's coverage probabilities are close to those of the QLR confidence set, which supports our theoretical results. We also report the coverage probabilities for the case $a_0 = 0.75$. Under this initial choice of level, the Wald confidence sets' coverage probabilities are closer to the nominal level in every case.

We note that the QLR confidence sets' coverage probabilities improve as we move from $n = 1{,}000$ to $n = 2{,}000$. This behavior is not so apparent for the Wald confidence sets.

[24] We use the dataset that is distributed with a Matlab package by Beresteanu, Molinari, and Wang (2009).

[25] The reported values are the average elapsed time to compute the support function and approximated boundary of the set estimator from simulated samples of size $n = 1,000$ or 2,000 drawn for $S = 2,000$ times. Computation was implemented by a code written in R (and partly in C) on a computer with Intel Core 2 Quad CPU 2.5 Ghz and 6GB memory.

Table 1: *Coverage Probabilities ($1-\alpha$=.95) of Wald and QLR Confidence Sets*

	Subsample Size				
n=1,000					
	$b=50$	$b=80$	$b=120$	$b=200$	$b=300$
Wald ($a_0=0.5$)	0.878	0.887	0.886	0.889	0.909
Wald ($a_0=0.75$)	0.957	0.962	0.953	0.956	0.956
QLR	0.851	0.880	0.872	0.931	0.912
n=2,000					
	$b=200$	$b=300$	$b=400$	$b=500$	$b=600$
Wald ($a_0=0.5$)	0.874	0.873	0.874	0.880	0.888
Wald ($a_0=0.75$)	0.937	0.925	0.925	0.930	0.928
QLR	0.861	0.882	0.905	0.950	0.933

Note: Empirical coverage probabilities of the Wald and QLR confidence sets under different values of subsample size b. The coverage probabilities of the QLR confidence set are taken from Chernozhukov, Hong, and Tamer (2004). Monte Carlo simulations $m=2,000$, subsample replications $B=2,000$, significance level $\alpha=0.05$.

6 Conclusion

In this paper, we introduce an inference 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. This yields inference tools that have the wide applicability of the criterion function approach and the computational tractability of the support function approach.

We consider the general case where the 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. This yields Wald-type inference methods, significantly extending recent work of BM and Bontemps, Magnac, and Maurin (2007), each of which studied special classes of econometric models. Specifically, given a compact convex set Θ_0 or a point θ_0, we present tests for set equality $H_0 : \Theta_I = \Theta_0$, set inclusion $H_0 : \Theta_0 \subseteq \Theta_I$, and point inclusion $H_0 : \theta_0 \in \Theta_I$.

The test for set equality can be inverted to construct a confidence collection that contains the identified set as an element, with a specified confidence level. This type of inference is as yet unavailable within CHT's framework. The test for set inclusion can be inverted to construct another confidence collection, containing each subset of the identified set as an element. Taking the union of the elements of this collection yields a confidence set that covers the identified set, comparable to CHT's confidence set. We provide a new, practical step-up algorithm for selecting the level t used to construct this confidence set. This removes the arbitrariness in the choice of t characterizing previous methods. The test for point inclusion can be inverted to construct a confidence set for each point in the identified set, comparable to methods of Imbens and Manski (2004), CHT, Romano and Shaikh (2008), and Andrews and Guggenberger (2009).

We also contribute to the literature on moment inequality models by establishing the asymptotic equivalence of our Wald statistic and CHT's QLR statistic. We show that this implies the asymptotic equivalence of the Wald confidence set and CHT's confidence set. This equivalence suggests that further investigation into the general relationship between these two approaches, beyond the moment inequality framework, is an interesting topic for future research.

Another interesting direction for further research is the development of Lagrange Multiplier (LM)-type analogs of the Wald-type statistics analyzed here. One may expect that under suitable conditions, LM- and Wald-type statistics may also be asymptotically equivalent in partially identified models, and that under further conditions, these may be asymptotically equivalent to QLR-type statistics. Obtaining these equivalence conditions is an interesting direction for future research.

For testing hypotheses and constructing confidence collections and confidence sets, we propose a general subsampling procedure. This procedure is valid pointwise, as we derive our results under a fixed probability measure. As Romano and Shaikh (2008, 2009) and Andrews and Guggenberger (2009) point out, however, establishing the uniform asymptotic validity of subsampling is important for partially identified models and is one of our future tasks.

A Mathematical Appendix

A.1 Consistency and Rate of Convergence of the Level Set Estimator

We summarize below CHT's consistency and the rate of convergence result. Assumption 2.3 (i) requires one-sided uniform convergence of Q_n to its population counterpart, which is slightly more general than usual uniform convergence $\sup_{\theta\in\Theta}|Q(\theta)-Q_n(\theta)|=o_p(1)$. Assumption 2.3 (ii) is one of the key conditions utilized by CHT, requiring the sample criterion function to approximate the population counterpart at $1/a_n$ rate over Θ_I. This condition ensures that their QLR-statistic $\sup_{\theta\in\Theta_I} a_nQ_n(\theta)$ is nondegenerate. Assumption 2.3 (iii) requires the existence of a polynomial function in the distance from Θ_I, which stochastically minorizes (bounds from below) the sample criterion function in a neighborhood of the identified set. It is then immediate from CHT's Theorem 3.1 that the following results hold.

Theorem A.1 (Consistency and Convergence rate): *Let t be a positive finite constant. Let $t_n = t\kappa_n$ where κ_n is a positive slowly increasing sequence such that $\kappa_n \to \infty$ and $\kappa_n/a_n = o_p(1)$. Suppose Assumptions 2.1, 2.2, and 2.3 (i), (ii) hold. Then, with probability approaching 1, $\hat{\Theta}_n(t) \subseteq \Theta_I$ and $\Theta_I \subseteq \hat{\Theta}_n(t_n)$. Furthermore, $d_H(\hat{\Theta}_n(t_n),\Theta_I) = o_p(1)$. Suppose, in addition, Assumption 2.3 (iii) holds. Then, $r_n d_H(\hat{\Theta}_n(t_n),\Theta_I) = O_p(1)$ with $r_n = (a_n/\max\{1,\kappa_n\})^{1/\gamma}$.*

For the proof, see CHT's Theorem 3.1.

A.2 Proof of Lemma 3.1, 3.2 , and Lemma 3.3

Definition A.1 (Level boundedness): *The function $f:\mathbb{R}^d\to\bar{\mathbb{R}}$ is* level-bounded *if the level sets $\{x : f(x)\le\alpha\}$ are bounded for any $\alpha\in\mathbb{R}$.*

If a function $f:\Omega\times\mathbb{R}^d\to\bar{\mathbb{R}}$ is such that $f(\omega,\cdot)$ is level bounded for all $\omega\in F\in\mathfrak{F}$, $P(F)=1$, then we say f is level bounded almost surely ($a.s.$).

Proof of Lemma 3.1. Note that Assumption 2.1 ensures that ζ_n is proper. In addition, the compactness of Θ and the assumption that $Q_n(\omega,\theta)=\infty$ $a.s.$ for $\theta\notin\Theta$ ensure that ζ_n is level-bounded almost surely. For each $\omega\in\{\omega:\zeta_n \text{ is lsc}\}$, we have

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

where the second equivalence follows from the compactness of $\hat{\Theta}_n(t)$, which is implied by the lower semicontinuity and the level-boundedness of ζ_n, and the last equivalence follows from the properness and the lower semicontinuity of ζ_n and the compactness of $K_{u,p}\cap\Theta$. □

Proof of Lemma 3.2. Note first that, under our assumptions, $\tilde{\zeta}_n$ inherits the almost sure properness, lower semicontinuity, and level-boundedness from ζ_n. For any $0 < \epsilon < \bar{\epsilon}$, let $D^*_{n,\epsilon} := \{(\theta, \lambda) \in R_{u,p} : \tilde{\zeta}_n(\theta, \lambda) < \inf_{R_{u,p}} \tilde{\zeta}_n(\theta, \lambda) + \epsilon\}$ and $D^*_\epsilon := \liminf_{n\to\infty} D^*_{n,\epsilon}$. By hypothesis, D^*_ϵ is nonempty. For a given $\delta > 0$, let $D^*_{\epsilon,\delta}$ be an open δ-envelope of D^*_ϵ defined by $D^*_{\epsilon,\delta} := \{(\theta, \lambda) : d((\theta, \lambda), D^*_\epsilon) < \delta\}$.

For any $\delta > 0$, $R_{u,p} \cap D^*_{\epsilon,\delta} \neq \varnothing$ implies that there exists $N_\epsilon \in \mathbb{N}$ such that $R_{n,u,p} \cap D^*_{\epsilon,\delta} \neq \varnothing$ for all $n \geq N_\epsilon$ as $R_{n,u,p} \to R_{u,p}$ in the Painlevé-Kuratowski sense (Theorem 4.5, Rockafellar and Wets, 2005). For $n \geq N_\epsilon$, let $E_{n,\epsilon} := \arg\min_{R_{n,u,p}\cap D^*_{\epsilon,\delta}} \tilde{\zeta}_n(\theta, \lambda)$. Let $D_n := \arg\min_{R_{n,u,p}} \tilde{\zeta}_n(\theta, \lambda)$. As $E_n \neq \varnothing$, $E_{n,\epsilon} \subseteq D^*_{\epsilon,\delta}$, and $E_{n,\epsilon} \subseteq D_n$, we have $D_n \cap D^*_{\epsilon,\delta} \neq \varnothing$ for all $n \geq N_\epsilon$ and $\delta > 0$.

Now suppose that the conclusion of the lemma does not hold. Then, there exists a subsequence $\{(\tilde{\zeta}_{n_k}, R_{n_k}), k = 1, 2, \cdots\}$ such that

$$P\left(\left|\inf_{R_{n_k,u,p}} \tilde{\zeta}_{n_k}(\theta, \lambda) - \inf_{R_{u,p}} \tilde{\zeta}_{n_k}(\theta, \lambda)\right| \geq 2\epsilon\right) > 0.$$

for all k. Then, along this subsequence, we have $P(D_{n_k} \cap D^*_{n_k,\epsilon} = \varnothing) > 0$. This implies $D_{n_k} \cap D^*_{\epsilon,\delta} = \varnothing$ for all k with positive probability, which is a contradiction. □

Proof of Lemma 3.3. Let $\epsilon > 0$ be arbitrary. For each $(u, p) \in \mathbb{R} \times \mathbb{S}^{d-1}$, take $L_{u,p}$ to be a compact set such that $D^*_\epsilon \subseteq L_{u,p}$. This is possible by the equi-inf-compactness. Now take $\tilde{R}_{u,p} = R_{u,p} \cap L_{u,p}$. Then, by construction,

$$P\left(\left|\inf_{\tilde{R}_{u,p}} \tilde{\zeta}_n(\theta, \lambda) - \inf_{R_{u,p}} \tilde{\zeta}_n(\theta, \lambda)\right| \geq \epsilon\right) \leq \epsilon \tag{A.1}$$

for sufficiently large n. Given the discussions preceding this lemma, we have

$$\begin{aligned} \mathcal{Z}_n(p, t) < u &\Leftrightarrow \inf_{R_{n,u,p}} \zeta_n(\theta + \lambda/a_n^{1/\gamma}) > t \\ &\Leftrightarrow \inf_{R_{n,u,p}} \tilde{\zeta}_n(\theta, \lambda) > t. \end{aligned}$$

Since this holds for any finite m-tuple $\{(u_j, p_j)\}_{j=1}^m$, we have

$$\begin{aligned} &P(\mathcal{Z}_n(p_1, t) < u_1, \cdots, \mathcal{Z}_n(p_m, t) < u_m) \\ &\qquad = P\left(\inf_{R_{n,u_1,p_1}} \tilde{\zeta}_n(\theta, \lambda) > t, \cdots, \inf_{R_{n,u_m,p_m}} \tilde{\zeta}_n(\theta, \lambda) > t\right). \end{aligned} \tag{A.2}$$

Note that

$$
\begin{aligned}
P\bigg(&\inf_{\tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon, \cdots, \inf_{\tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon\bigg) \\
&\le P\left(\max_{1\le j\le m}\left|\inf_{\tilde{R}_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) - \inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda)\right| \ge \epsilon/2\right) \\
&\quad + P\left(\inf_{R_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon/2, \cdots, \inf_{R_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon/2\right) \\
&\le P\left(\max_{1\le j\le m}\left|\inf_{\tilde{R}_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) - \inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda)\right| \ge \epsilon/2\right) \qquad \text{(A.3)} \\
&\quad + P\left(\max_{1\le j\le m}\left|\inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) - \inf_{R_{n,u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda)\right| \ge \epsilon/2\right) \\
&\quad + P\left(\inf_{R_{n,u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{R_{n,u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t\right),
\end{aligned}
$$

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 $G = (t, \infty)^m$.

Lemma 3.2 and (A.1) ensure that the first two terms on the right hand side of (A.3) become arbitrarily small as n gets large. Therefore,

$$
\begin{aligned}
\liminf_{n\to\infty} P\bigg(&\inf_{\tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon, \cdots, \inf_{\tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon\bigg) \\
&\le \liminf_{n\to\infty} P\left(\inf_{R_{n,u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{R_{n,u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t\right).
\end{aligned}
$$

By letting $\epsilon \downarrow 0$, we obtain

$$
\begin{aligned}
\liminf_{n\to\infty} P\bigg(&\inf_{\tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{\tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t\bigg) \\
&\le \liminf_{n\to\infty} P\left(\inf_{R_{n,u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{R_{n,u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t\right) \\
&= \liminf_{n\to\infty} P(\mathcal{Z}_n(p_1,t) < u_1, \cdots, \mathcal{Z}_n(p_m,t) < u_m),
\end{aligned}
$$

where the last equality follows from Eq. (A.2). □

A.3 Proof of Theorem 3.1 (i), (ii) and Auxiliary Lemmas

Our first goal in this sections is to show that the stochastic process $\mathcal{Z}(\cdot, t)$ given in Eq. (3.3) in Theorem 3.1 satisfies $\{\omega : \mathcal{Z}(p,t) < u\} = \{\omega : \inf_{\tilde{R}_{u,p}} \tilde{\zeta}(\theta,\lambda) > t\}$ for any $u, p \in \mathbb{R} \times \mathbb{S}^{d-1}$. For this, we need to show the almost sure upper semicontinuity of the map $g : \theta \mapsto s(p, \hat{\Lambda}(t,\theta))$. In the following, we introduce a regularity condition for the criterion function and two lemmas that are useful for establishing the desired result. We then prove Theorem 3.1 (i) and (ii).

Definition A.2 (Level-boundedness for parametric optimization): *A function* $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ *with values* $f(x,u)$ *is level-bounded in* x *locally uniformly in* u *if for each* $\bar{u} \in \mathbb{R}^m$ *and* $\alpha \in \mathbb{R}$ *there is a neighborhood* $V \in \mathcal{N}(\bar{u})$ *along with a bounded set* $B \subset \mathbb{R}^n$ *such that* $\{x | f(x,u) \leq \alpha\} \subset B$ *for all* $u \in V$*; or equivalently, there is a neighborhood* $V \in \mathcal{N}(\bar{u})$ *such that the set* $\{(x,u) | u \in V, f(x,u) \leq \alpha\}$ *is bounded in* $\mathbb{R}^n \times \mathbb{R}^m$.

Lemma A.1: *Consider*

$$\psi(u) := \inf_x f(x,u)$$

in the case of a proper, lsc function $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ *such that* $f(x,u)$ *is level bounded in* x *locally uniformly in* u*. Then, the function* ψ *is proper and lsc on* $\mathbb{R}^m$.

Proof. See Theorem 1.17 in Rockafellar and Wets (2005) □

Lemma A.2: *Suppose that* $\tilde{\zeta}_n(\theta,\lambda)$ *satisfies the conditions of Theorem 3.1. For each* $t \in \mathbb{R}_+$ *and* $p \in \mathbb{S}$*, let* g *be a stochastic process defined by* $g : \theta \mapsto s\left(p, \hat{\Lambda}(t,\theta)\right)$*. Then, there is a representation of* g*, which is upper semicontinuous (usc) almost surely.*

Proof. First, let $\delta_A : \mathbb{R}^d \to \bar{\mathbb{R}}_+$ be the optimization theory indicator function that takes 0 if $x \in A$ and ∞ otherwise. For each θ, let $h(\theta,\lambda) := -\langle p,\lambda\rangle + \delta_{\hat{\Lambda}(t,\theta)}(\lambda)$ and $\tilde{g}(\theta) := \inf_\lambda h(\theta,\lambda)$. As $g(\theta) = -\tilde{g}(\theta)$, it suffices to show the lower semicontinuity of $\tilde{g}(\theta)$ for the conclusion of the lemma[26]. For establishing the lower semicontinuity of $\tilde{g}$, we make use of Lemma A.1 by taking $\psi = \tilde{g}$, $f = h$, and $(x,u) = (\theta,\lambda)$. Below, we show that h is almost surely proper, lsc, and level bounded in λ locally uniformly in θ.

By our hypothesis, $\hat{\Lambda}(t,\theta)$ is nonempty *a.s.* for any $\theta \in \partial\Theta_I$ and $t \in \mathbb{R}_+$. Therefore, $\delta_{\hat{\Lambda}(t,\theta)}$ is proper, which implies that h is proper. In the following, using Skrokhod representation, we take a version of $\tilde{\zeta}_n$ that is epiconverging almost surely to a version of $\tilde{\zeta}$ that are defined on some common probability space. This is possible since the space of lsc convex functions is Polish (Geyer 2003, p.5-6.). The almost sure epiconvergence of lsc functions $\{\tilde{\zeta}_n, n \geq 1\}$ implies that $\tilde{\zeta}$ is lsc *a.s.* (Attouch, 1984, Theorem 2.1). Therefore, the level set $\hat{\Lambda}(t,\theta)$ of the lsc function $\tilde{\zeta}(\theta,\cdot)$ is closed *a.s.* Note that $-\langle p,\lambda\rangle$ is continuous and $\delta_{\hat{\Lambda}(t,\theta)}$ is lsc by the closedness of $\hat{\Lambda}(t,\theta)$. So, h is lsc *a.s.*

For each $p \in \mathbb{S}^{d-1}$ and $\bar{\theta} \in H(p,\Theta_I)$, let $\mathcal{N}(\bar{\theta})$ be a collection of neighborhoods at $\bar{\theta}$. Let $\alpha \in \mathbb{R}$. Take $\theta \in V \in \mathcal{N}(\bar{\theta})$. Define the set

$$C := \{\lambda : \tilde{\zeta}(\theta,\lambda) \leq t, \quad \langle p,\lambda\rangle \geq -\alpha\}.$$

The fact that $\tilde{\zeta}_n$ is equi-inf-compact implies that $\tilde{\zeta}$ is level bounded (Exercise 7.32 (b) in Rockafellar and Wets (2005)), and therefore $\hat{\Lambda}(t,\theta)$ is bounded *a.s.* As $C \subseteq \hat{\Lambda}(t,\theta)$, C is bounded *a.s.* Now, we can rewrite

$$\begin{aligned} C &= \{\lambda : \delta_{\Lambda(t,\theta)}(\lambda) = 0, \quad -\langle p,\lambda\rangle \leq \alpha\} \\ &= \{\lambda : h(\theta,\lambda) \leq \alpha\}. \end{aligned}$$

Therefore, $h(\theta,\lambda)$ is level bounded in λ locally uniformly in θ. By Lemma A.1, $\tilde{g}(\theta)$ is lsc

[26] We follow the convention that $\sup_{x\in C} f(x) = -\infty$ if C is an empty set.

almost surely. ☐

Given the results above, we fist prove the statement of Theorem 3.1 (ii).

Proof of Theorem 3.1 (ii). For each $(u,p) \in \mathbb{R} \times \mathbb{S}^{d-1}$, take $L_{u,p}$ to be a compact set such that $\hat{\Lambda}(t,\theta) \subset L_{u,p}$ *a.s.* For a given $\theta \in H(p,\Theta_I)$, it is straightforward to show

$$s(p,\hat{\Lambda}(t,\theta)) < u \Leftrightarrow \inf_{\lambda \in K_{u,p} \cap L_{u,p}} \tilde{\zeta}(\theta,\lambda) > t, \tag{A.4}$$

using an argument similar to the proof of Lemma 3.1. By the compactness of $H(p,\Theta_I)$ and Lemma A.2,

$$\begin{aligned} \mathcal{Z}(p,t) < u \Leftrightarrow & \sup_{\theta \in H(p,\Theta_I)} s(p,\hat{\Lambda}(t,\theta)) < u \\ & \Leftrightarrow s(p,\hat{\Lambda}(t,\theta)) < u, \quad \forall \theta \in H(p,\Theta_I). \end{aligned} \tag{A.5}$$

Combining Eqs. (A.4) and (A.5), we obtain

$$\mathcal{Z}(p,t) < u \Leftrightarrow \inf_{(\theta,\lambda) \in \tilde{R}_{u,p}} \tilde{\zeta}(\theta,\lambda) > t,$$

where $\tilde{R}_{u,p} = H(p,\Theta_I) \times (K_{u,p} \cap L_{u,p})$. Therefore, for any finite m-tuple $\{(u_j,p_j)\}_{j=1}^m$,

$$\begin{aligned} &\{\omega : \mathcal{Z}(p_1,t) < u_1, \cdots, \mathcal{Z}(p_m,t) < u_m\} \\ &\qquad = \left\{\omega : \inf_{(\theta,\lambda) \in \tilde{R}_{u_1,p_1}} \tilde{\zeta}(\theta,\lambda) > t, \cdots, \inf_{(\theta,\lambda) \in \tilde{R}_{u_m,p_m}} \tilde{\zeta}(\theta,\lambda) > t \right\}. \end{aligned}$$

Take probability both sides. Then, the conclusion of Theorem 3.1 (ii) follows. ☐

Proof of Theorem 3.1 (i). Consider a finite m-tuple $\{(u_j,p_j)\}_{j=1}^m$. Since $\tilde{\zeta}_n \overset{e.d.}{\to} \tilde{\zeta}$, for any $\{(u_j,p_j)\}_{j=1}^m$, we have

$$\begin{aligned} & P\left(\inf_{(\theta,\lambda) \in \tilde{R}_{u_1,p_1}} \tilde{\zeta}(\theta,\lambda) > t, \cdots, \inf_{(\theta,\lambda) \in \tilde{R}_{u_m,p_m}} \tilde{\zeta}(\theta,\lambda) > t \right) \\ & \quad \le \liminf_{n\to\infty} P\left(\inf_{(\theta,\lambda) \in \tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{(\theta,\lambda) \in \tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t \right). \end{aligned}$$

This result, together with Lemma 3.3 and Theorem 3.1 (ii) proved above, implies that

$$P\left(\mathcal{Z}(p_1,t) < u_1, \cdots, \mathcal{Z}(p_m,t) < u_m\right) \le \liminf_{n\to\infty} P\left(\mathcal{Z}_n(p_1,t) < u_1, \cdots, \mathcal{Z}_n(p_m,t) < u_m\right).$$

By the portmanteau theorem, the process $\mathcal{Z}_n(\cdot,t)$ weakly converges to $\mathcal{Z}(\cdot,t)$ in finite dimension. This completes the proof of part (i). ☐

A.4 Proof of Theorem 3.1 (iii)

For establishing $\mathcal{Z}_n(\cdot,t) \overset{u.d.}{\to} \mathcal{Z}(\cdot,t)$, we make use of the three lemmas below. Lemmas A.3 and A.4 will be used to show the tightness of the sequence $\{\mathcal{Z}_n(\cdot,t), n=1,2,\cdots\}$. Lemma A.3 states that, for the tightness, it suffices to show the stochastic equicontinuity of the process, and Lemma A.4 gives a sufficient condition for the stochastic equicontinuity.

Lemma A.3 (Tightness Characterization): *Let $\mathbb{E}$ be a metric space. A sequence of stochastic processes $\{\xi_n(x), n \geq 1\}$ is tight in $(\mathcal{C}(\mathbb{E}), d_{\mathcal{C}(\mathbb{E})})$ if and only if $\xi_n(x) = O_p(1)$ for all $x \in \mathbb{E}$ and the stochastic equicontinuity holds. That is, For every $\epsilon, \eta > 0$ there exists random $\Delta_n(\epsilon,\eta)$ and a constant $N_{\epsilon,\eta}$ such that for $n \geq N_{\epsilon,\eta}$, $P(|\Delta_n(\epsilon,\eta)| > \epsilon) < \eta$ and for each $y \in \mathbb{E}$, there is an open set $V(y,\epsilon,\eta)$ containing θ with*

$$\sup_{x \in V(y,\epsilon,\eta)} |\xi_n(x) - \xi_n(y)| \leq \Delta_n(\epsilon,\eta), \quad n \geq N_{\epsilon,\eta}.$$

Proof. See Newey (1991). □

Lemma A.4: *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 $x, y \in \mathbb{E}$, $|\xi_n(x) - \xi_n(y)| \leq B_n h(\|x-y\|)$, then $\{\xi_n\}$ is stochastically equicontinuous.*

Proof. The result immediately follows from Assumption 3A and Corollary 2.2 in Newey (1991). □

The lemmas above imply that showing that $\{\mathcal{Z}_n(\cdot,t), n \geq 1\}$ satisfies the regularity conditions in Lemma A.4 suffices for the desired result. For this, we make use of the following definition and the lemma.

Definition A.3 (Strict Continuity): *Let $\mathcal{S} \subseteq \mathbb{R}^d$. A function $f : \mathcal{S} \to \mathbb{R}$ is strictly continuous at $\bar{x} \in \mathcal{S}$ if $\bar{x} \in \mathcal{S}^o$ and if the Lipshitz modulus, $lipf(\bar{x}) := \limsup_{\substack{x,x' \to \bar{x} \\ x \neq x'}} \frac{|f(x') - f(x)|}{\|x' - x\|}$, is finite. A function is strictly continuous on $\mathcal{S}$ if it is strictly continuous at every point in $\mathcal{S}$.*

We say a function $f : \Omega \times \mathcal{S} \to \bar{\mathbb{R}}$ is strictly continuous on $\mathcal{S}$ almost surely ($a.s.$) if $f(\omega, \cdot)$ is strictly continuous for all $\omega \in F \in \mathfrak{F}, P(F) = 1$. For single-valued mappings, the strict continuity is equivalent to local Lipschitz property, i.e., the function is Lipschitz on a neighborhood of each point (Rockafellar and Wets, 2005).

Lemma A.5 (Extended Mean Value Theorem): *Suppose f is convex and strictly continuous on an open convex set $O \subset \mathbb{R}^d$, and let x_0 and x_1 be points of O. Then there exist $x_\tau = (1-\tau)x_0 + \tau x_1, \tau \in (0,1)$ and $v \in \mathbb{R}^d$ satisfying*

$$f(x_1) - f(x_0) = \langle v, x_1 - x_0 \rangle, \quad v \in \partial f(x_\tau),$$

where $\partial f(x)$ is the subdifferential of f at x, defined by

$$\partial f(x) := \{v \in \mathbb{R}^d : f(y) \geq f(x) + \langle v, y - x \rangle, \quad \forall y \in \mathbb{R}^d\}.$$

Proof. See Rockafellar and Wets (2005), Theorem 10.48. □

Proof of Theorem 3.1 (iii). We first show the required conditions for Lemma A.4 using an expansion of the support function based on Lemma A.5. In the following, we take $\omega \in F \in \mathfrak{F}$, $P(F) = 1$, for which $\tilde{\zeta}_n(\omega, \theta, \lambda)$ is lsc and convex. Under our assumptions, Θ_I is a compact convex set, and $\hat{\Theta}_n(t)$ is a compact convex set almost surely. For each bounded closed set, its support function is Lipschitz (Theorem F.1. in Molchanov (2005)). This implies that $s(p, \Theta_I)$ is strictly continuous, and $s(p, \hat{\Theta}_n(t))$ is strictly continuous *a.s*. Furthermore, the support function of a compact set is sublinear, and therefore it is convex (Molchanov, 2005, p.421). This further implies that $s(p, \Theta_I)$ is convex, and $s(p, \hat{\Theta}_n(t))$ 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 Lemma A.5, 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

$$s(p, \hat{\Theta}_n(t)) - s(q, \hat{\Theta}_n(t)) = \langle \hat{v}_n, p - q \rangle \tag{A.6}$$

$$s(p, \Theta_I) - s(q, \Theta_I) = \langle w, p - q \rangle \tag{A.7}$$

For any compact convex set F, the subdifferential $\partial s(p, F)$ of the support function at p coincides with its support set $H(p, F)$. Therefore, $\partial s(\bar{p}_n, \hat{\Theta}_n(t)) = H(\bar{p}_n, \hat{\Theta}_n(t))$ and $\partial s(\bar{p}, \Theta_I) = H(\bar{p}, \Theta_I)$. So, we can write

$$\mathcal{Z}_n(p, t) - \mathcal{Z}_n(q, t) = a_n^{1/\gamma} \langle \hat{v}_n - w, p - q \rangle \tag{A.8}$$

for some $\hat{v}_n \in H(\bar{p}_n, \hat{\Theta}_n(t))$ and $w \in H(\bar{p}, \Theta_I)$.

Note that, Assumption 2.3 (ii) implies $\mathcal{Z}_n(p, t) = O_p(1)$ 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-Schwartz inequality to (A.8), we obtain

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

Now, we apply Lemma A.4 with $B_n = a_n^{1/\gamma} \|\hat{v}_n - w\|$ and $h(x) = x$. Then, $\{\mathcal{Z}_n(\cdot, t), n \geq 1\}$ is stochastically equicontinuous. Further, we apply Lemma A.3 to conclude that $\{\mathcal{Z}_n(\cdot, t), n \geq 1\}$ 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 $\mathcal{Z}_n(\cdot, t) \stackrel{u.d.}{\to} \mathcal{Z}(\cdot, t)$.

Since ζ_n is convex *a.s.* in a neighborhood of Θ_I, its lower contour sets (in the neighborhood of Θ_I) are convex almost surely. Thus the Hörmander's embedding theorem implies

$$a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) = \|\mathcal{Z}_n(\cdot, t)\|_{\mathcal{C}(\mathbb{S})}.$$

By the continuous mapping theorem, we get $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \stackrel{d}{\to} \|\mathcal{Z}(\cdot, t)\|_{\mathcal{C}(\mathbb{S})}$. Similarly, by Theorem 2.3, the directed Hausdorff distance satisfies $a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) = \sup_{p \in \mathbb{S}^{d-1}} \{-\mathcal{Z}_n(p, t)\}_+$. The continuous mapping theorem implies $a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \stackrel{d}{\to} \sup_{p \in \mathbb{S}^{d-1}} \{-\mathcal{Z}(p, t)\}_+$ □

A.5 Proof of Theorems and Corollaries in Section 3.2

Proof of Theorem 3.2. By Theorem 2.2, $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) = a_n^{1/\gamma} \|s(\cdot, \hat{\Theta}_n(t)) - s(\cdot, \Theta_I)\|_{\mathcal{C}(\mathbb{S}^{d-1})}$. Here, $s(\cdot, \hat{\Theta}_n(t))$ is a random element that takes values in a normed linear space $(\mathcal{C}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathcal{C}(\mathbb{S}^{d-1})})$. Let $F_n(x,t)$ be the cdf of $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I)$. Note that Theorem 3.1 ensures F_n converges weakly to F. Now apply Politis, Romano, and Wolf's (1999) Theorem 2.5.2. with $\tau_n = a_n^{1/\gamma}$, $\hat{\theta}_n = s(\cdot, \hat{\Theta}_n(t))$, and $\theta(P) = s(\cdot, \Theta_I)$. Then, all the results follow. □

Proof of Corollary 3.1. (i) First, the consistency of $\hat{c}_{n,b,1-\alpha}(t)$ follows from Theorem 3.2 and Lemma 11.2.1 in Lehmann and Romano (2005). Under the null hypothesis, $T_n(t)$ converges in distribution to $F(x,t)$, and by the result above, $\hat{c}_{n,b,1-\alpha}(t) = c_{1-\alpha}(t) + o_p(1)$. Then, by Corollary 11.2.3 in Lehmann and Romano (2005), $\lim_{n\to\infty} P(T_n(t) \le \hat{c}_{n,b,1-\alpha}(t)) = F(c_{1-\alpha}(t), t) = 1-\alpha$.

(ii) The proof of part (ii) is very similar to the proof of Theorem 2.2 in BM. □

Proof of Theorem 3.3. The first part follows from the equivalence

$$\Theta_I \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t) \quad \Leftrightarrow \quad a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \le \hat{c}_{n,b,1-\alpha}(t).$$

and Theorem 3.2 (iii). Note that for any compact convex set $K \in \mathcal{K}_c$ and $\epsilon > 0$, we have

$$K \oplus B_\epsilon = K^\epsilon,$$

where B_ϵ is a closed ball of radius ϵ centered at the origin, and K^ϵ is a closed ϵ-envelope of K. The rest of the proof is very similar to the proof of Theorem 2.4 in BM.

(iii) Note that, for any $t \in \mathbb{R}_+$, $\Theta_I \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t)$ implies $\Theta_I \subseteq \hat{\Psi}_{n,b,1-\alpha}(t)$, but the converse is not necessarily true. Therefore, by part (i),

$$\lim_{n\to\infty} P\left(\Theta_I \subseteq \hat{\Psi}_{n,b,1-\alpha}(t)\right) \ge \lim_{n\to\infty} P\left(\Theta_I \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t)\right) = 1-\alpha.$$

By part (ii), $\hat{\Psi}_{n,b,1-\alpha}(t) = \hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}(t)}(t)$. Now, the conclusion follows. □

Proof of Theorem 3.4. For each $t \in \mathbb{R}_+$, let $U_{n,b}^{\rightarrow}(x,t) := N_{n,b}^{-1} \sum_{k=1}^{N_{n,b}} 1_{\left\{a_b^{1/b} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) \le x\right\}}$. Suppose, for any $\epsilon > 0$,

$$a_b^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \hat{\Theta}_{n,b,k}(t)) = a_b^{1/\gamma} \sup_{p \in \mathbb{S}^{d-1}} \left\{ s(p, \hat{\Theta}_n(t)) - s(p, \hat{\Theta}_{n,b,k}(t)) \right\}_+ \le x$$

and

$$a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) = a_b^{1/\gamma} \sup_{p \in \mathbb{S}^{d-1}} \left| s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I) \right| \le \epsilon.$$

Then, $a_b^{1/\gamma}(s(p, \Theta_I) - s(p, \hat{\Theta}_{n,b,k}(t))) \le x + \epsilon$ for all $p \in \mathbb{S}^{d-1}$. This further implies

$$a_b^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) = a_b^{1/\gamma} \sup_{p \in \mathbb{S}^{d-1}} \left\{ s(p, \Theta_I) - s(p, \hat{\Theta}_{n,b,k}(t)) \right\}_+ \le x + \epsilon.$$

Let $E_{n,b}(t,\epsilon) := \{\omega : a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \epsilon\}$. Then, the arguments above ensure

$$\hat{F}_{n,b}^{\rightarrow}(x,t) 1_{\{E_{n,b}(t,\epsilon)\}} \leq U_{n,b}^{\rightarrow}(x+\epsilon, t). \tag{A.9}$$

Now, suppose

$$a_b^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) \leq x - \epsilon$$

and

$$a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \epsilon.$$

Then, we have

$$\begin{aligned} x &\geq a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) + a_b^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) \\ &\geq a_b^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \Theta_I) + a_b^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) \\ &\geq a_b^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \hat{\Theta}_{n,b,k}(t)). \end{aligned}$$

Therefore, we obtain

$$U_{n,b}^{\rightarrow}(x-\epsilon, t) 1_{\{E_{n,b}(t,\epsilon)\}} \leq \hat{F}_{n,b}^{\rightarrow}(x,t) 1_{\{E_{n,b}(t,\epsilon)\}}. \tag{A.10}$$

Since (A.9) and (A.10) hold for any $\epsilon > 0$ and $E_{n,b}$ has probability tending to one, we have

$$U_{n,b}^{\rightarrow}(x-\epsilon, t) \leq \hat{F}_{n,b}^{\rightarrow}(x,t) \leq U_{n,b}^{\rightarrow}(x+\epsilon, t), \tag{A.11}$$

with probability tending to 1 for any $\epsilon > 0$.

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

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

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

The proofs of (ii) and (iii) are very similar to those of Theorem 2.2.1 (ii) and (iii) in Politis, Romano, and Wolf (1999). □

Proof of Corollary 3.2. (i) As before, the consistency of $\hat{c}_{n,b,1-\alpha}^{\rightarrow}(t)$ follows from Lemma 11.2.1 in Lehmann and Romano (2005). Under the null, we have $T_n^{\rightarrow}(t) \leq a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t))$, and $\vec{d}_H(\Theta_I, \hat{\Theta}_n(t))$ converges in distribution to $F^{\rightarrow}(x,t)$. By the results above and by Corollary 11.2.3 in Lehmann and Romano (2005), we have

$$\begin{aligned} \limsup_{n\to\infty} P(T_n^{\rightarrow}(t) > \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t)) &\leq \lim_{n\to\infty} P(a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) > \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t)) \\ &= 1 - F(c_{1-\alpha}^{\rightarrow}(t), t) = \alpha. \end{aligned}$$

The proof of part (ii) is similar to the proof of Corollary 3.1 (ii). □

Proof of Theorem 3.5. The first part follows from the equivalence

$$\Theta_I \in \hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t) \quad \Leftrightarrow \quad a_n^{1/\gamma}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \leq \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t).$$

and Theorem 3.4 (iii). The proof of (ii) is similar to the proof of Theorem 3.3 (ii). The proof of (iii) is similar to the proof of Proposition 2.7 in BM. □

Proof of Lemma 3.4. First, $c^{\rightarrow}_{1-\alpha}(t^*_{1-\alpha}) = 0$ follows from the definition of $t^*_{1-\alpha}$. For the conclusion of the lemma, it suffices to show that $P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \leq x\right)$ is non-decreasing in t for each x. As this is a distributoinal property of the process $\mathcal{Z}(p,t)$, it suffices to show that the statement above holds for the following representation:

$$-\mathcal{Z}(p,t) = -\sup_{\theta\in H(p,\Theta_I)} \sup_{\lambda\in\{\lambda:\tilde{\zeta}(\theta,\lambda)\leq t\}} \langle p, \lambda\rangle.$$

As $\{\lambda : \tilde{\zeta}(\theta,\lambda) \leq t\} \subseteq \{\lambda : \tilde{\zeta}(\theta,\lambda) \leq t'\}$ for any $0 \leq t < t' \leq t^*_{1-\alpha}$ 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, $P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \leq x\right)$ is non-decreasing in $t \in [0, t^*_{1-\alpha}]$ for each x. □

We use the following lemma to prove Theorem 3.6.

LEMMA A.6: *Suppose the conditions of Theorem 3.6 hold. Then, for any $\alpha \in (0,1)$ and $0 \leq t < t' \leq t^*_{1-\alpha}$, $c^{\rightarrow}_{1-\alpha}(t) - c^{\rightarrow}_{1-\alpha}(t') = \mu(t') - \mu(t)$.*

Proof of Lemma A.6. First, $c^{\rightarrow}_{1-\alpha}(t)$ can be written as

$$\begin{aligned} c^{\rightarrow}_{1-\alpha}(t) &= \inf\left\{x : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \leq x\right) \geq 1-\alpha\right\} \\ &= \inf\left\{x : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{\mu(t') - \mu(t) - \mu(t') - \mathcal{Z}^*(p)\}_+ \leq x\right) \geq 1-\alpha\right\}. \end{aligned} \tag{A.12}$$

Let $\Delta(t,t') := \mu(t') - \mu(t)$. Then, for any $x \geq \Delta(t,t')$, we have

$$\begin{aligned} &P\left(\sup_{p\in\mathbb{S}^{d-1}}\{\mu(t') - \mu(t) - \mu(t') - \mathcal{Z}^*(p)\}_+ \leq x\right) \\ &= P\left(\sup_{p\in\mathbb{S}^{d-1}}\{\Delta(t,t') - \mathcal{Z}(p,t')\}_+ \leq x\right) \\ &= P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t')\}_+ \leq x - \Delta(t,t')\right). \end{aligned} \tag{A.13}$$

Substituting Eq. (A.13) into Eq. (A.12) yields

$$\begin{aligned} c^{\rightarrow}_{1-\alpha}(t) &= \inf\left\{x : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t')\}_+ \leq x - \Delta(t,t')\right) \geq 1-\alpha\right\} \\ &= c^{\rightarrow}_{1-\alpha}(t') + \Delta(t,t'). \end{aligned}$$

□

Proof of Theorem 3.6. By Theorem 2.2,

$$
\begin{aligned}
a_n^{1/\gamma} d_H\left(\hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}^{\rightarrow}(t)}(t), \hat{\Theta}_n(t^*_{1-\alpha})\right) &= a_n^{1/\gamma} \sup_{p\in\mathbb{S}^{d-1}} |s(p, \hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}^{\rightarrow}(t)}(t)) - s(p, \hat{\Theta}_n(t^*_{1-\alpha}))| \\
&= a_n^{1/\gamma} \sup_{p\in\mathbb{S}^{d-1}} |s(p, \hat{\Theta}_n(t)) + \hat{\epsilon}_{n,b,1-\alpha}^{\rightarrow}(t) - s(p, \hat{\Theta}_n(t^*_{1-\alpha}))| \\
&= \sup_{p\in\mathbb{S}^{d-1}} |a_n^{1/\gamma}[s(p, \hat{\Theta}_n(t)) - s(p, \hat{\Theta}_n(t^*_{1-\alpha}))] + \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t)| \\
&= \sup_{p\in\mathbb{S}^{d-1}} |a_n^{1/\gamma}[s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I)] \\
&\qquad - a_n^{1/\gamma}[s(p, \hat{\Theta}_n(t^*_{1-\alpha})) - s(p, \Theta_I)] + \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t)| \\
&= \sup_{p\in\mathbb{S}^{d-1}} |\mathcal{Z}_n(p,t) - \mathcal{Z}_n(p, t^*_{1-\alpha}) + c_{1-\alpha}^{\rightarrow}(t) + o_p(1)| \\
&\overset{(1)}{=} \sup_{p\in\mathbb{S}^{d-1}} |\mu(t) - \mu(t^*_{1-\alpha}) - (c_{1-\alpha}^{\rightarrow}(t^*_{1-\alpha}) - c_{1-\alpha}^{\rightarrow}(t)) + o_p(1)| \\
&= o_p(1),
\end{aligned}
$$

where we used the fact that $c_{1-\alpha}^{\rightarrow}(t^*_{1-\alpha}) = 0$ in equality (1), and the last equality follows from Lemma A.6. □

Proof of Corollary 3.3. The result immediately follows from Theorem 3.6 and the triangle inequality:

$$
\begin{aligned}
d_H\left(\hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}^{\rightarrow}(t)}(t), \hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}^{\rightarrow}(t')}(t'))\right) &\leq \\
d_H\left(\hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}^{\rightarrow}(t)}(t), \hat{\Theta}_n(t^*_{1-\alpha}))\right) &+ d_H\left(\hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}^{\rightarrow}(t')}(t'), \hat{\Theta}_n(t^*_{1-\alpha}))\right). \quad \square
\end{aligned}
$$

Proof of Theorem 3.7. Let l be the smallest random index for which there is a false rejection. Then, there is $\Theta_0 \in S_l^c \cap \mathcal{K}_{cI}$ such that

$$
a_n^{1/\gamma} \vec{d}_H(\Theta_0, \hat{\Theta}_n(t_l)) > \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t_l).
$$

As $\Theta_0 \subseteq \Theta_I$, this implies

$$
\begin{aligned}
a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t_l)) &> \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t_l) \\
&\Rightarrow a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n^{\hat{c}_{n,b,1-\alpha}^{\rightarrow}(t_l)/a_n^{1/\gamma}}(t_l)) > 0 \\
&\Rightarrow a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t^*_{1-\alpha})) + o_p(1) > 0,
\end{aligned}
$$

where the last result follows from Theorem 3.6. Therefore,

$$
\lim_{n\to\infty} FWER \leq P\left(a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t^*_{1-\alpha})) + o_p(1) > 0\right) = \alpha. \square
$$

A.6 Proofs for Theorems and Corollaries in Section 3.4

The following result immediately follows from the main theorem.

Lemma A.7: *Suppose conditions of Theorem 3.1 hold. Suppose that the weak epilimit $\tilde{\zeta}$ is such that for any $p, q \in S_{\theta_0}$,*

$$\sup_{\lambda \in \{\lambda : \tilde{\zeta}(\theta_0, \lambda) \le t\}} \langle p, \lambda \rangle = \sup_{\lambda \in \{\lambda : \tilde{\zeta}(\theta_0, \lambda) \le t\}} \langle q, \lambda \rangle \tag{A.14}$$

for any $\omega \in \Omega$. Then, for each $t \in \mathbb{R}_+$,

(i) if $\theta_0 \in \partial\Theta_I$, $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{d} \{-\mathcal{Z}(p_0, t)\}_+$;

(ii) if $\theta_0 \in \Theta_I^o$, $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{p} 0$;

(iii) if $\theta_0 \notin \Theta_I$, $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{p} +\infty$.

Proof of Lemma A.7. The proof is similar to that of Proposition 16 in Bontemps, Magnac, and Maurin (2007). Let $\mathcal{A}_n(p, t) := a_n^{1/\gamma}[\langle p, \theta_0 \rangle - s(p, \hat{\Theta}_n(t))]$ and $\mathcal{A}(p, t) := a_n^{1/\gamma}[\langle p, \theta_0 \rangle - s(p, \Theta_I)]$. First, we show that $\mathcal{A}_n(\hat{p}_n, t) - \mathcal{A}_n(p_0, t) = o_p(1)$. Note that $\mathcal{A}_n(\hat{p}_n, t) - \mathcal{A}_n(p_0, t)$ is bounded from below by 0, as $\hat{p}_n$ maximizes $\mathcal{A}_n(\cdot, t)$. From above, we have

$$\begin{aligned} 0 \le \mathcal{A}_n(\hat{p}_n, t) - \mathcal{A}_n(p_0, t) &= \mathcal{A}_n(\hat{p}_n, t) - \mathcal{A}(\hat{p}_n, t) + \mathcal{A}(\hat{p}_n, t) - \mathcal{A}(p_0, t) + \mathcal{A}(p_0, t) - \mathcal{A}_n(p_0, t) \\ &= -\mathcal{Z}_n(\hat{p}_n, t) + \mathcal{A}(\hat{p}_n, t) - \mathcal{A}(p_0, t) + \mathcal{Z}_n(p_0, t) \\ &\le \mathcal{Z}_n(p_0, t) - \mathcal{Z}_n(\hat{p}_n, t) \\ &= \mathcal{Z}(p_0, t) - \mathcal{Z}(\hat{p}_n, t) + u_p(1), \end{aligned}$$

where the inequality holds because $\mathcal{A}(\hat{p}_n, t) - \mathcal{A}(p_0, t) \le 0$ by the construction of p_0, and $u_p(1)$ represents a term that is uniformly converging in probability to 0.

Note that $\{\hat{p}_n\}$ has a subsequence $\{\hat{p}_{n_l}\}$ that converges in probability to some $p' \in S_{\theta_0}$. Then, by the stochastic equicontinuity, $\mathcal{Z}(\hat{p}_{n_l}, t) \xrightarrow{p} \mathcal{Z}(p', t)$. Note that Eq. (3.3) and (A.14) imply $\mathcal{Z}(p', t) = \mathcal{Z}(p_0, t)$, $P - a.s.$ for any $p', p_0 \in S_{\theta_0}$. Therefore, along any subsequence, $\mathcal{Z}(\hat{p}_{n_l}, t)$ converges in probability to $\mathcal{Z}(p_0, t)$. This implies $\mathcal{Z}(p_0, t) - \mathcal{Z}(\hat{p}_n, t) = o_p(1)$.

(i) Suppose $\theta_0 \in \partial\Theta_I$, then $\langle p_0, \theta_0 \rangle = s(p_0, \Theta_I)$. Therefore, $\mathcal{A}_n(p_0, t) = -\mathcal{Z}_n(p_0, t)$. Now,

$$\begin{aligned} T_{n,\theta_0}(t) &= \{\mathcal{A}_n(\hat{p}_n, t)\}_+ \\ &= \{\mathcal{A}_n(p_0, t)\}_+ + o_p(1) \\ &= \{-\mathcal{Z}_n(p_0, t)\}_+ + o_p(1). \end{aligned}$$

By Theorem 3.1 (i) and the continuous mapping theorem, $T_{n,\theta_0}(t) \xrightarrow{d} \{-\mathcal{Z}(p_0, t)\}_+$. (ii) Suppose $\theta \in \Theta_I^o$, then $T_{n,\theta_0}(t) = \{-\mathcal{Z}_n(p_0, t) + a_n^{1/\gamma}[\langle p_0, \theta_0 \rangle - s(p_0, \Theta_I)]\}_+ + o_p(1)$ and $a_n^{1/\gamma}[\langle p_0, \theta_0 \rangle - s(p_0, \Theta_I)] \to -\infty$. Therefore, $T_{n,\theta_0}(t) \xrightarrow{p} 0$. (iii) Suppose $\theta \notin \Theta_I$, then $T_{n,\theta_0}(t) = \{-\mathcal{Z}_n(p_0, t) + a_n^{1/\gamma}[\langle p_0, \theta_0 \rangle - s(p_0, \Theta_I)]\}_+ + o_p(1)$ and $a_n^{1/\gamma}[\langle p_0, \theta_0 \rangle - s(p_0, \Theta_I)] \to \infty$. Therefore, $T_{n,\theta_0}(t) \xrightarrow{p} \infty$. □

Proof of Theorem 3.8. The proof is very similar to the proof of Theorem 3.4, and therefore it is omitted. □

Proof of Corollary 3.4. (i) First $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(p_0, t) + o_p(1)$. Then, the consistency of $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t)$ follows by applying Lemma 11.2.1 in Lehmann and Romano (2005). If $\theta_0 \in \partial\Theta_I$, $T^{\rightarrow}_{n,\theta_0}(t)$ converges in distribution to $F^{\rightarrow}(x, p, t)$ by Lemma A.7 (i), and by the result above, $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t) = c^{\rightarrow}_{1-\alpha}(p_0, t) + o_p(1)$. Then, by Corollary 11.2.3 in Lehmann and Romano (2005), $\lim_{n\to\infty} P(T^{\rightarrow}_{n,\theta_0}(t) \le \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n,t)) = F^{\rightarrow}(c^{\rightarrow}_{1-\alpha}(p_0,t),t) = 1-\alpha$. If $\theta \in \Theta_I^o$, $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{p} 0$ by Lemma A.7 (ii), and therefore $\lim_{n\to\infty} P(T^{\rightarrow}_{n,\theta_0}(t) \le \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n,t)) \ge \lim_{n\to\infty} P(\{-\mathcal{Z}_n(\hat{p}_n,t)\}_+ \le \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n,t)) = F^{\rightarrow}(c^{\rightarrow}_{1-\alpha}(p_0,t),t) = 1-\alpha$.

(ii) The proof of part (ii) is a direct consequence of Lemma A.7 (iii) and $\hat{c}_{n,b,1-\alpha}(\hat{p}_n,t) = O_p(1)$. □

Proof of Theorem 3.9. The result simply follows from the equivalence

$$\theta_0 \in \tilde{\Psi}^{\rightarrow}_{n,b,1-\alpha}(t) \quad \Leftrightarrow \quad T^{\rightarrow}_{n,\theta_0} \le \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n,t).$$

and Corollary 3.4 (i). □

Proof of Theorem 3.10. First, we can rewrite the local power as

$$\begin{aligned}\pi_{n,b,t}(\theta_n) &= P\left(a_n^{1/\gamma}\{\langle \hat{p}_n,\theta_n\rangle - s(\hat{p}_n,\hat{\Theta}_n(t))\}_+ > \hat{c}_{n,b,1-\alpha}(\hat{p}_n,t)\right)\\ &= P\left(a_n^{1/\gamma}\{s(\hat{p}_n,\Theta_I) - s(\hat{p}_n,\hat{\Theta}_n(t)) + \langle \hat{p}_n,\lambda\rangle/a_n^{1/\gamma}\}_+ > \hat{c}_{n,b,1-\alpha}(\hat{p}_n,t)\right)\\ &= P\left(\{-\mathcal{Z}_n(\hat{p}_n,t) + \langle \hat{p}_n,\lambda\rangle\}_+ > \hat{c}_{n,b,1-\alpha}(\hat{p}_n,t)\right).\end{aligned}$$

By the fact that $\langle \hat{p}_n,\lambda\rangle = \langle p_0,\lambda\rangle + o_p(1) = h + o_p(1)$ and the stochastic equicontinuity of $\mathcal{Z}(\cdot,t)$ and $\hat{c}_{n,b,1-\alpha}(\cdot,t)$, we obtain

$$\pi_{n,b,t}(\theta_n) = P(\{-\mathcal{Z}_n(p_0,t) + h + o_p(1)\}_+ > c^{\rightarrow}_{1-\alpha}(p_0,t) + o_p(1)).$$

Note that $\lim_{n\to\infty} \pi_{n,b,t}(\theta_n) \ge P(\{-\mathcal{Z}(p_0,t)\}_+ > c^{\rightarrow}_{1-\alpha}(p_0,t)) = \alpha$. So, the test has the asymptotic local unbiasedness. Furthermore, using the fact that $\{-x\}_+ > \epsilon \Leftrightarrow x < -\epsilon$ for any $\epsilon \ge 0$, we can write

$$\pi_{n,b,t}(\theta_n) = P(\mathcal{Z}_n(p_0,t) < h - c^{\rightarrow}_{1-\alpha}(p_0,t)) + o(1).$$

By the second duality (Lemma 3.3) and the weak epiconvergence

$$\begin{aligned}\liminf_{n\to\infty} P(\mathcal{Z}_n(p_0,t) < h - c^{\rightarrow}_{1-\alpha}(p_0,t)) &\ge \liminf_{n\to\infty} P\left(\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}_n(\theta,\lambda) > t\right)\\ &\ge P\left(\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}(\theta,\lambda) > t\right),\end{aligned}$$

where $\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} = \{\theta_0\}\times(K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} \cap L_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0})$ with $L_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}$ properly chosen. □

Proof of Corollary 3.5. The proof of part (i) is very similar to the proof of Corollary 3.2 (i). For part (ii), we make use of the reverse triangle inequality:

$$\vec{d}_H(\{\theta_0\},\Theta_I) - \vec{d}_H(\Theta_I,\hat{\Theta}_n(t)) \le \vec{d}_H(\{\theta_0\},\hat{\Theta}_n(t)).$$

Scaling both sides by $a_n^{1/\gamma}$ and taking $\inf_{\Theta_0}$ both sides give

$$a_n^{1/\gamma}\left(\inf_{\theta_0\in\Theta_0}\vec{d}_H(\{\theta_0\},\Theta_I)-\vec{d}_H(\Theta_I,\hat{\Theta}_n(t))\right)\leq\inf_{\theta_0\in\Theta_0}T_{n,\theta_0}^{\rightarrow}(t). \tag{A.15}$$

Since $\hat{c}_{n,b,1-\alpha}^{\rightarrow}(t)=c_{1-\alpha}^{\rightarrow}(t)+o_p(1)$ and $\vec{d}_H(\Theta_I,\hat{\Theta}_n(t))=o_p(1)$, we have

$$\begin{aligned}\lim_{n\to\infty}P\bigg(\hat{c}_{n,b,1-\alpha}^{\rightarrow}(t)&<a_n^{1/\gamma}\left(\inf_{\theta_0\in\Theta_0}\vec{d}_H(\{\theta_0\},\Theta_I)-\vec{d}_H(\Theta_I,\hat{\Theta}_n(t))\right)\bigg)\\&=\lim_{n\to\infty}P\left(\frac{c_{1-\alpha}^{\rightarrow}(t)+o_p(1)}{a_n^{1/\gamma}}<\inf_{\theta_0\in\Theta_0}\vec{d}_H(\{\theta_0\},\Theta_I)+o_p(1)\right)\\&=1,\end{aligned}$$

since $\inf_{\theta_0\in\Theta_0}\vec{d}_H(\{\theta_0\},\Theta_I)>0$ under the alternative. By the inequality A.15, the conclusion follows. □

A.7 Proof of Theorems and Corollaries in Section 4

Knight (1999) provides a result that links the weak finite dimensional limit and the weak epilimit, which we summarize below as a lemma. For this result, we introduce the following definition.

Definition A.4 (Stochastic Equi-lowersemicontinuity): *A sequence of random lsc functions $\{\xi_n, n\geq 1\}$ on $\mathbb{R}^d$ is stochastically equi-lowersemicontinuous (e-lsc) if for each bounded set B, $\epsilon>0$ and $\delta>0$, there exist $x_1,\cdots,x_m\in B$ and open neighborhoods $V(x_1),\cdots,V(x_m)$ of $x_1,\cdots,x_m$ such that*

$$B\subset\bigcup_{i=1}^m V(x_i)$$

and

$$\limsup_{n\to\infty}P\left(\bigcup_{i=1}^m\left\{\inf_{y\in V(x_i)}\xi_n(y)\leq\min\{\epsilon^{-1},\xi_n(x_i)-\epsilon\}\right\}\right)<\delta.$$

Lemma A.8 (Knight, 1999, Theorem 2): *Let $\{\xi_n, n\geq 1\}$ be a stochastically e-lsc sequence of functions and ξ be a random lsc function. Then $\xi_n\overset{f.d.}{\to}\xi$ if and only if $\xi_n\overset{e.d.}{\to}\xi$.*

Proof of Theorem 4.1. Assumption 2.1 immediately follows from Assumption 4.1. Assumption 2.2 immediately follows from Assumption 4.2. For the consistency of the level set estimator with the choice of finite nonnegative constant t, we additionally need to show Assumptions 2.3 (i), (ii), and 2.4 (i).

By Assumption 4.3 (i-a,b), we can write

$$
\begin{aligned}
&P\left(\sup_{\Theta}\left|\varphi(\hat{E}_n(m_{j,\theta},\hat{W}_n(\theta))-\varphi(E(m_{j,\theta}),W(\theta))\right|>\epsilon\right)\\
&\quad\le P\Bigg(\sup_{\Theta}\left|\varphi(\hat{E}_n(m_{j,\theta},\hat{W}_n(\theta))-\varphi(E(m_{j,\theta}),\hat{W}_n(\theta))\right|\\
&\qquad\quad+\sup_{\Theta}\left|\varphi(E(m_{j,\theta}),\hat{W}_n(\theta))-\varphi(E(m_{j,\theta}),W(\theta))\right|>\epsilon\Bigg)\\
&\quad\le P\Bigg(\sup_{\Theta}\left|\varphi(\hat{E}_n(m_{j,\theta},\hat{W}_n(\theta))-\varphi(E(m_{j,\theta}),\hat{W}_n(\theta))\right|>\epsilon/2\Bigg)\\
&\qquad\quad+P\Bigg(\sup_{\Theta}\left|\varphi(E(m_{j,\theta}),\hat{W}_n(\theta))-\varphi(E(m_{j,\theta}),W(\theta))\right|>\epsilon/2\Bigg)\\
&\quad\le P\left(\sup_{\Theta}L_1h_1\left(\left\|\hat{E}_n(m_{j,\theta})-E(m_{j,\theta})\right\|\right)>\epsilon/2\right)\\
&\qquad\quad+P\left(\sup_{\Theta}L_2h_2\left(\max_{i,j}\left|\hat{W}_{n,ij}(\theta)-W_{ij}(\theta)\right|\right)>\epsilon/2\right)\\
&\quad\le P\left(\sup_{\Theta}\left\|\hat{E}_n(m_{j,\theta})-E(m_{j,\theta})\right\|>L_1^{-1}h_1^{-1}(\epsilon/2)\right)\\
&\qquad\quad+P\left(\sup_{\Theta}\max_{i,j}\left|\hat{W}_{n,ij}(\theta)-W_{ij}(\theta)\right|>L_2^{-1}h_2^{-1}(\epsilon/2)\right)\\
&\quad\le\epsilon
\end{aligned}
$$

for n sufficiently large. Therefore, Assumptions 2.3 (i) holds. In the following, we take $a_n = n^{\gamma/2}$. First, this choice of a_n and the P-donsker property ensure that $\sup_{\Theta_I} a_n Q_n(\theta) = O_p(1)$. Therefore, Assumptions 2.3 (ii) holds. Now, let $\eta > 0$ be such that $\sup_{\Theta_I}\max_{i,j}\hat{W}_{n,ij}(\theta) \le \eta < \infty$, $wp \to 1$. We can write

$$
\begin{aligned}
n^{\gamma/2}Q_n(\theta) &\le \varphi(\mathbb{G}_n(m_\theta)+\sqrt{n}E(m_\theta),\hat{W}_n(\theta))\\
&\le \varphi\left(\mathbb{G}_n(m_\theta)+\sqrt{n}E(m_\theta),\sup_{\Theta_I}\max_{i,j}|\hat{W}_{n,ij}(\theta)|I_J\right)\\
&\le \varphi\left(O_p(1)-\sqrt{n}C_3\min\{d(\theta,\Theta\setminus\Theta_I),\bar{\epsilon}\},\eta I_J\right)
\end{aligned}
$$

uniformly over Θ_I $wp \to 1$ by Assumption 4.3 (ii-d). We thus have $Q_n(\theta) = 0$ on $\Theta_I^{-\epsilon_n}$ with $\epsilon_n = O_p(1/\sqrt{n})$, and this ensures Assumption 2.4 (i).

For the rate of convergence of the set estimator, we additionally need to show Assumptions 2.3 (iii) and 2.4 (ii). For this, we closely follow CHT's proof of Theorem 4.2. Take $\eta' > 0$ such that $\inf_{\Theta_I}\min_{i,j}|\hat{W}_{n,ij}(\theta)| \ge \eta'$, $wp \to 1$. First, we write

$$
\begin{aligned}
n^{\gamma/2}Q_n(\theta) &= \varphi(\mathbb{G}_n(m_\theta)+\sqrt{n}E(m_\theta),\hat{W}_n(\theta))\\
&\ge C_2\|\mathbb{G}_n(m_\theta)+\sqrt{n}E(m_\theta)\|_+^\gamma\\
&\ge C_2\|\sqrt{n}E(m_\theta)\|_+^\gamma\frac{\|\mathbb{G}_n(m_\theta)+\sqrt{n}E(m_\theta)\|_+^\gamma}{\|\sqrt{n}E(m_\theta)\|_+^\gamma}.
\end{aligned}
$$

Now, by Assumption 4.3 (ii-b), we have $\|\sqrt{n}E(m_\theta)\|_+^\gamma \ge C_1 n^{\gamma/2}\min\{d(\theta,\Theta_I),\delta\}^\gamma$ on Θ for some $C_1 > 0$ and $\delta > 0$. Therefore, for any $\epsilon > 0$, we can choose $(\kappa_\epsilon, n_\epsilon)$ so that for any $n \geq n_\epsilon$ with probability at least $1 - \epsilon$,

$$n^{\gamma/2} Q_n(\theta) \geq C_2 C_1 n^{\gamma/2} \min\{d(\theta, \Theta_I), \delta\}^\gamma,$$

uniformly over $\{\theta \in \Theta : d(\theta, \Theta_I) \geq (\kappa_\epsilon / n^{\gamma/2})^{1/\gamma}\}$, which follows by $\|y + x\|_+ / \|x\|_+ \to 1$ as $\|x\|_+ \to \infty$ for any $y \in \mathbb{R}^J$ and by $\sup_{\Theta_I} \|\mathbb{G}_n(m_\theta)\| = O_p(1)$ by the P-Donsker property.

Note that the lower semicontinuity $\tilde{\zeta}_n$ follows the continuity in θ of φ, m, and W. The convexity of $Q_n(\theta)$ in a neighborhood of Θ_I directly follows from Assumption 4.4 (ii). Now, we show the weak epiconvergence of $\tilde{\zeta}_n$. First, $\mathbb{G}_n m_\theta \overset{u.d.}{\to} \mathbb{G}(\theta)$ implies $\mathbb{G}_n(m_\theta) \overset{f.d.}{\to} \mathbb{G}(\theta)$. Together with Assumption 4.4 (i-a,b), this implies $\mathcal{M}_n(\theta, \lambda) \overset{f.d.}{\to} \mathcal{M}(\theta, \lambda)$. We also have $\hat{W}_n(\theta + \lambda / a_n^{1/\gamma}) \overset{p}{\to} W(\theta)$. Therefore, by the continuous mapping theorem, $\tilde{\zeta}_n = \varphi(\mathcal{M}_n(\theta, \lambda), \hat{W}_n(\theta + \lambda / a_n^{1/\gamma})) \overset{f.d.}{\to} \varphi(\mathcal{M}(\theta, \lambda), W(\theta)) = \tilde{\zeta}$. Now by Assumption 4.4 (iii), we can apply Lemma A.8 to conclude that $\tilde{\zeta}_n \overset{e.d.}{\to} \tilde{\zeta}$. □

Proof of Corollary 4.1. By Theorem 4.1, the conditions required for Theorem 3.1 hold. The weak convergence results immediately follow from Theorem 3.1, and the representation result follows from the fact that $\tilde{\zeta}(\theta, \lambda) = \varphi(\mathbb{G}(\theta) + \Pi(\theta)\lambda + \varsigma(\theta))$. □

Proof of Corollary 4.2. Let $s : \partial\Theta \times \mathbb{R}^d \to \mathbb{R}^{J(\theta)}$ be a vector-valued mapping whose j-th component is $s_j(\theta, \lambda) = 1\{\mathbb{G}_j(\theta) + \langle \Pi_j(\theta), \lambda \rangle > 0\}$. As the linear constraint qualification is satisfied, the solution λ^* to the minimization problem (4.3) satisfies the following Karush-Kuhn-Tucker (KKT) conditions with probability 1[27]:

$$\begin{aligned}
&p = 2\mu \Pi_{\mathcal{J}(\theta)}(\theta)' W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*) \\
&t \geq \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2 \\
&0 \leq \mu \\
&0 = \mu(\|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2 - t),
\end{aligned}$$

where μ is the Lagrange multiplier associated with the constraint in Eq. (4.3). By Assumption 4.5 (ii), the constraint in (4.3) binds, and the conditions above simplify to

$$p = 2\mu \Pi_{\mathcal{J}(\theta)}(\theta)' W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*) \tag{A.16}$$

$$t = \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2 \tag{A.17}$$

$$\mu > 0. \tag{A.18}$$

We can solve (A.16) to obtain

$$\begin{aligned}
&\big(W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)' W^{1/2}_{\mathcal{J}(\theta)}(\theta)\big)^{-1} W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta) p \\
&\qquad = 2\mu W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*).
\end{aligned} \tag{A.19}$$

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

[27] The constraint is non-differentiable only at finite number of points, and the probability of $\mathbb{G}(\theta)$ taking these values is 0.

obtain

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

where the second equality follows from (A.17). So, we obtain

$$2\mu = \|\mathcal{R}(p,\theta)\| t^{-1/2}. \tag{A.20}$$

Plugging this into (A.19) gives

$$W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*) = \frac{\mathcal{R}(p,\theta)}{\|\mathcal{R}(p,\theta)\|} t^{1/2}. \tag{A.21}$$

Substituting (A.20) and (A.21) into (A.16) yields

$$p = \Pi'_{\mathcal{J}(\theta)} W^{1/2}_{\mathcal{J}(\theta)} \mathcal{R}(p,\theta). \tag{A.22}$$

Now, we can use this result to obtain

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

where the fourth equality follows from the fact that $\mathcal{R}(p,\theta) = \mathcal{R}(p,\theta) \circ s(\theta,\lambda^*)$, and the fifth equality follows from (A.21).

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

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

Note that Eq. (A.22) implies that

$$p'p = p'\Pi_{\mathcal{J}(\theta)}(\theta)'(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p.$$

As p is in the unit sphere, $p'p = \|p\|^2 = 1$. 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

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

□

For Theorem 4.2, we require the following regularity conditions.

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

$$\lim_{n\to\infty} P\left(\sup_{\|p-q\|<\delta} \left| \sup_{R^-_{u,p}} \tilde{\zeta}_n(\theta,\lambda) - \sup_{R^-_{u,q}} \tilde{\zeta}_n(\theta,\lambda) \right| \geq \epsilon \right) \leq \epsilon,$$

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

Assumption A.1 (i) requires that the finite dimensional distribution of the supremum of $\tilde{\zeta}_n$ over a class of compact sets converges to that of $\tilde{\zeta}$. This is analogous to weak epiconvergence. We call this version "weak supconvergence" as it is close in spirit to Condition S.2 of CHT.

Proof of Theorem 4.2. First, by the hypothesis that $\tilde{\zeta}_n$ weakly supconverges to $\tilde{\zeta}$, $\mathcal{L}_n(\cdot,u) \xrightarrow{f.d.} \mathcal{L}(\cdot,u)$ where

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

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

$$\begin{aligned} \mathcal{C}(\theta,p,u) := \sup_{\lambda} &\|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta)+\Pi_{\mathcal{J}(\theta)}(\theta)\lambda)\|^2_+ \\ s.t. &\langle p,\lambda\rangle \leq u. \end{aligned} \tag{A.23}$$

Similar to the proof of Corollary 4.2, the solution λ^* of the problem above satisfies the following KKT conditions with probability 1.

$$\begin{aligned} \nu p &= 2\Pi_{\mathcal{J}(\theta)}(\theta)' W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta)+\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*) \\ \langle p,\lambda^*\rangle &\leq u \\ 0 &\leq \nu \\ 0 &= \nu(u - \langle p,\lambda^*\rangle), \end{aligned}$$

where ν is the Lagrange multiplier associated with the constraint in (A.23). By Assumption 4.5 (ii), the constraint in (A.23) binds, and the conditions above simplify to

$$\nu p = 2\Pi_{\mathcal{J}(\theta)}(\theta)' W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta)+\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*) \tag{A.24}$$

$$\langle p,\lambda^*\rangle = u \tag{A.25}$$

$$0 < \nu.$$

We can solve (A.24) to obtain

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

Taking squared norms both sides, we obtain

$$\begin{aligned}\nu^2\|\mathcal{R}(p,\theta)\|^2 &= 4\|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*)\|^2 \\ &= 4\mathcal{C}(\theta,p,u).\end{aligned} \tag{A.27}$$

Plugging in $\nu = 2\mathcal{C}(\theta,p,u)^{1/2}/\|\mathcal{R}(p,\theta)\|$ back to (A.24), we obtain

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

Now, substitute this into (A.25),

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

where the second equality follows from (A.26). Using (A.26) and the result above, the right hand side of (A.27) can be alternatively written as

$$\begin{aligned}&2\nu\left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle + \left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\right\rangle\right) \\ &= 2\nu\left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle + u\right).\end{aligned}$$

Therefore, from (A.27), we obtain

$$\begin{aligned}\nu &= 2\|\mathcal{R}(p,\theta)\|^{-1}\left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle + u\right) \\ &= 2\|\mathcal{R}(p,\theta)\|^{-1}\left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle + u\right)_+,\end{aligned}$$

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

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

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

Proof of Corollary 4.3. We first analyze the Wald statistic $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}_n(p,t) + t^{1/2}\}_+^2$. By Corollary 4.2, the distributional limit $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t) + t^{1/2}\}_+^2$ of this statistic can be represented as

$$
\begin{aligned}
&\sup_{p\in\mathbb{S}^{d-1}}\left\{-\sup_{\theta\in H(p,\Theta_I)}-\left\langle(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p,\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle\right\}_+^2\\
&=\sup_{p\in\mathbb{S}^{d-1}}\left\{\inf_{\theta\in H(p,\Theta_I)}\left\langle(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p,\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle\right\}_+^2\\
&=\sup_{p\in\mathbb{S}^{d-1}}\left\langle\left(\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))'\right)^{-1}\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))p,\mathbb{G}_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\right\rangle_+^2\\
&=\mathbf{Z},
\end{aligned}
$$

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

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

by Theorem 4.2 and the continuous mapping theorem. By Theorem 4.2, this limit can be represented as

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

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 that, we can write

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

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

A.8 Proof of Theorems in Section 5.1

The following lemma is often useful to identify the weak epilimit of a sequence of stochastic processes.

LEMMA A.9: *Let $\Gamma(\mathbb{R}^d)$ be the space of convex lsc functions on $\mathbb{R}^d$ that are proper and have effective domains with nonempty interiors (or equivalently are finite on an open set). Suppose that $\{\xi_n, n \geq 1\}$ is a sequence in $\Gamma(\mathbb{R}^d)$ and let $\mathbb{Q}$ be a countable dense subset of $\mathbb{R}^d$. If $\xi_n \overset{f.d.}{\to} \xi$ on $\mathbb{Q}$ where $P(\xi \in \Gamma(\mathbb{R}^d)) = 1$, then $\xi_n \overset{e.d.}{\to} \xi$.*

Proof. See Lemma 3.1. in Geyer (2003). □

Proof of Theorem 5.1. Let

$$\begin{aligned}\tilde{\zeta}_n(\theta, \lambda) &= nQ_n(\theta + \lambda/\sqrt{n}) \\ &= (\sqrt{n}(\hat{E}_n(X_{1i}) - \theta_1) - \lambda + \sqrt{n}(\theta_1 - \theta))_+^2 \\ &\quad + (\sqrt{n}(\theta - \theta_2) + \lambda - \sqrt{n}(\hat{E}_n(X_{2i}) - \theta_2))_+^2 + \infty \times 1_{\theta \notin \Theta}.\end{aligned}$$

This function is convex in (θ, λ), lsc, and has an effective domain with nonempty interior. Under our hypothesis, the finite dimensional limit of $\tilde{\zeta}_n(\theta, \lambda)$ is

$$\tilde{\zeta}(\theta, \lambda) = (Z_1 - \lambda + \varsigma_1(\theta))_+^2 + (\varsigma_2(\theta) + \lambda - Z_2)_+^2,$$

where $(Z_1, Z_2)' \sim N(0, \Omega)$ and

$$\varsigma_1(\theta) = \begin{cases} \infty & \theta < \theta_1 \\ 0 & \theta = \theta_1 \\ -\infty & \theta > \theta_1 \end{cases}, \qquad \varsigma_2(\theta) = \begin{cases} \infty & \theta > \theta_2 \\ 0 & \theta = \theta_2 \\ -\infty & \theta < \theta_2 \end{cases}.$$

This function is convex and lsc, and finite on an open interval (θ_1, θ_2), and $\tilde{\zeta}_n(\theta, \lambda) \overset{f.d.}{\to} \tilde{\zeta}(\theta, \lambda)$. Therefore, Lemma A.9 is applicable. Thus, the weak epi-limit coincides with the finite dimensional limit.

Using the representation result in Corollary 4.2, we can derive a closed form for $\mathcal{Z}$. For example, when $p = -1$ and $\theta \in H(-1, \theta) = \{\theta_1\}$, we have $\mathcal{J}(\theta_1) = 1$, $\mathcal{R}(-1, \theta) = 1$. Therefore,

$$\mathcal{Z}(-1, t) = t^{1/2} - Z_1.$$

Similarly,

$$\mathcal{Z}(1, t) = t^{1/2} + Z_2.$$

Therefore, the limiting process $\mathcal{Z}(p, t)$ has mean $t^{1/2}\iota$ and covariance $E[(\mathcal{Z}(-1,t) - t^{1/2})(\mathcal{Z}(-1,t) - t^{1/2})] = \Omega_{11}$, $E[(\mathcal{Z}(1,t) - t^{1/2})(\mathcal{Z}(1,t) - t^{1/2})] = \Omega_{22}$, and $E[(\mathcal{Z}(-1,t) - t^{1/2})(\mathcal{Z}(1,t) - t^{1/2})] = -\Omega_{12}$. By Corollary 4.1,

$$\sqrt{n} d_H(\hat{\Theta}_n(t), \Theta_I) \overset{d}{\to} \max\{|\mathcal{Z}(-1,t)|, |\mathcal{Z}(1,t)|\}$$

and

$$\sqrt{n} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \overset{d}{\to} \max\{-\mathcal{Z}(-1,t)_+, -\mathcal{Z}(1,t)_+\}. \square$$

Proof of Theorem 5.1. The result for $\mathcal{W}_n$ follows directly from Theorem 5.1. The results for $\mathcal{QLR}_n$ and $\tilde{\mathcal{W}}_n$ are due to Chernozhukov, Hong, and Tamer (2004) and BM respectively. $\square$

References

ANDERSEN, P. K., AND R. D. GILL (1982): "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10(4), 1100–1120.

ANDREWS, D. W. K. (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67(6), 1341–1383.

ANDREWS, D. W. K., S. BERRY, AND P. JIA (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," Working Paper, Yale University.

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 (2009): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, forthcoming.

ATTOUCH, H. (1986): *Variational Convergence for Functions and Operators (Applicable Mathematics Series)*. Pitman (Advanced Publishing Program), Boston, MA.

BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

BEER, G. A. (1993): *Topologies on Closed and Closed Convex Sets (Mathematics and Its Applications)*. Kluwer Academic Publishers, Dordrecht.

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

BERESTEANU, A., F. MOLINARI, AND E. Y. WANG (2009): "SVRV for Matlab: Beta Version: Set Valued Random Variables Estimation and Inference," Duke University.

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2007): "Set Identified Linear Models," Working Paper, Toulouse School of Economics.

BUGNI, F. A. (2009a): "Bootstrap Inference in Partially Identified Models," Working Paper, Duke University.

——— (2009b): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, forthcoming.

CANAY, I. A. (2009): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Working Paper, Northwestern University.

CHERNOFF, H. (1954): "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 25(3), 573–578.

CHERNOZHUKOV, V. (2005): "Extremal Quantile Regression," *Annals of Statistics*, 33(2), 806–839.

Chernozhukov, V., and H. Hong (2001): "Likelihood Inference with Density Jump," Working Paper, Massachusetts Institute of Technology.

——— (2004): "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models," *Econometrica*, 72(5), 1445–1480.

Chernozhukov, V., H. Hong, and E. Tamer (2004): "Inference on Parameter Sets in Econometric Models," Working Paper, Massachusetts Institute of Technology.

——— (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

Fan, Y., and S. S. Park (2007): "Confidence Sets for Some Partially Identified Parameters," Working Paper, Vanderbilt University.

Galichon, A., and M. Henry (2007): "Dilation Bootstrap: A Natural Approach to Inference in Incomplete Models," Columbia University, Working Paper.

——— (2009): "A test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics*, 152(2), 186 – 196, Nonparametric and Robust Methods in Econometrics.

Geyer, C. J. (1994): "On the Asymptotics of Constrained M-Estimation," *The Annals of Statistics*, 22(4), 1993–2010.

——— (2003): "On the Asymptotics of Convex Stochastic Optimization," Unpublished Manuscript, University of Minnesota.

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

Hahn, J., and G. Ridder (2009): "Partial Identification and Confidence Intervals," Working Paper, University of California, Los Angeles.

Han, C., and P. C. B. Phillips (2006): "GMM with Many Moment Conditions," *Econometrica*, 74(1), 147–192.

Hörmander, L. (1955): "Sur la Fonction D'Appui des Ensembles Convexes Dans un Espace Localement Convexes," *Arkiv fur Matematik*, 3, 181–186.

Horowitz, J. L., and C. F. Manski (1998): "Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations," *Journal of Econometrics*, 84(1), 37–58.

——— (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95(449), 77–84.

Imbens, G. W., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Kaido, H., and H. White (2008): "A Two-Stage Procedure for Partially Identified Models," Discussion Paper, University of California San Diego.

——— (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," Working Paper, University of Tronto.

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

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

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*. Berlin: Springer.

Moon, H. R., and F. Schorfheide (2009): "Bayesian and Frequentist Inference in Partially Identified Models," Working Paper, University of Southern California.

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

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 36. New York: Elsevier.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2005): "Moment Inequalities and Their Application," Working Paper, Harvard University.

Pflug, G. C. (1992): "Asymptotic Dominance and Confidence for Solutions of Stochastic Programs," *Czechoslovak Journal for Operations Ressearch*, 1, 21–30.

——— (1995): "Asymptotic Stochastic Programs," *Mathematics of Operations Research*, 20(4), 769–789.

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

Rockafellar, R. T., and R. J.-B. Wets (2005): *Variational Analysis*. Berlin: Springer, 2nd edn.

Romano, J. P., and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, Special Issue in Honor of T. W. Anderson.

——— (2009): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, forthcoming.

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.

Salinetti, G., and R. J.-B. Wets (1986): "On the Convergence in Distribution of Measurable Multifunctions (Random Sets), Normal Integrands, Stochastic Processes and Stochastic Infima," *Mathematics of Operations Research*, 11(3), 385–419.

Santos, A. (2007): "Inference in Nonparametric Instrumental Variables with Partial Identification," Working Paper, University of California San Diego.

Stoye, J. (2009): "More on Confidence Regions for Partially Identified Parameters," *Econometrica*, 77(4), 1299–1315.

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

White, H. (2001): *Asymptotic Theory for Econometricians, Revised Edition*. San Diego: Academic Press.

Woutersen, T. (2006): "A Simple Way to Calculate Confidence Intervals for Partially Identified," Working Paper, Johns Hopkins University.

Yildiz, N. (2009): "Consistency of Plug-In Estimators of Upper Contour Consistency of Plug-In Estimators of Upper Contour and Level Sets," Working Paper, University of Rochester.