A Dual Approach to Inference for Partially Identified Econometric Models*

Hiroaki Kaido
Boston University
Email: hkaido@bu.edu

February, 2012

Abstract

This paper considers inference for the set $\Theta_I$ 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 establishes a dual relationship between the level-set estimator and its support function and shows that the properly normalized (scaled and centered) support function provides an alternative Wald-type inference method to conduct tests regarding the identified set and a point $\theta_0$ in the identified set. These tests can be inverted to obtain confidence sets for $\Theta_I$ and $\theta_0$. For econometric models that involve finitely many moment inequalities, we show that our Wald-type statistic is asymptotically equivalent to CHT’s QLR statistic under some regularity conditions.

JEL Classification: C12
Keywords: Partial Identification, Criterion Function, Support Function

*I am deeply grateful to Halbert White for his guidance and constant encouragement. I thank Graham Elliott, Ben Gillen, Patrik Guggenberger, Ivana Komunjer, Francesca Molinari, Andres Santos, Yixiao Sun, Allan Timmermann, and seminar participants at UCSD, University of Arizona, UCLA, Penn State, BU, Cornell, University of Michigan, Brown, IUPUI, and Northwestern 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 $\Theta_I$, 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 $\Theta_I$ 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 $\Theta_I$ directly. This estimate can then be used to conduct inference for $\Theta_I$. This is an attractive alternative if the boundary of the identified set is easily estimable. Much of the literature has studied the case where $\Theta_I$ is a closed interval (e.g. Horowitz and Manski, 1998, 2000; Manski, 2003; Imbens and Manski, 2004). Recent studies extend this approach to the case where $\Theta_I$ is a multi-dimensional compact convex set (Beresteanu and Molinari, 2008 (BM); Bontemps, Magnac, and Maurin, forthcoming). When $\Theta_I$ is compact and convex, its support function provides a tractable representation by summarizing the location of the supporting hyperplanes of $\Theta_I$.

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 Tamer,

\(^1\)Our analysis applies to the convex hull of the identified set if it is nonconvex.
2002; Magnac and Maurin, 2008), a class of discrete choice models Pakes (2010), consumer demand models with unobserved heterogeneity (Blundell, Kristensen, and Matzkin, 2011), and an asset pricing model in incomplete markets (Kaido and White, 2009). Following CHT, our estimator of $\Theta_I$ 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.

Our approach is particularly well suited to conducting hypothesis tests and constructing 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 make various types of inference for $\Theta_I$ and a point in $\Theta_I$. For example, as done in BM, one may test whether a given set $\Theta_0$ is a subset of $\Theta_I$. This test can be inverted to construct a confidence set that covers the identified set with at least some prescribed confidence level. This confidence set is comparable to CHT’s confidence sets, constructed by inverting their QLR statistic. Further, one may test whether $\Theta_I$ includes a specific point, i.e., $H_0 : \theta_0 \in \Theta_I$. A test statistic is constructed using the estimated support function. We contribute to the literature by establishing the asymptotic distribution of this statistic. Specifically, our asymptotic distribution result generally holds even if the identified set has kinks and extends the result of Bontemps, Magnac, and Maurin (forthcoming). This test 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), Romano and Shaikh (2008), and Andrews and Guggenberger (2009).

The construction of confidence sets by inverting the normalized support function was first proposed by BM for the case where $\Theta_I$ is a linear transformation of the Aumann expectation of set-valued random variables. Bontemps, Magnac, and Maurin (forthcoming) consider a confidence set for a point in the identified set, when $\Theta_I$ is characterized by incomplete linear moment restrictions. Our analysis further contributes by extending these results to the general case where $\Theta_I$ 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; Han and Phillips, 2006). Our analysis shows that weak epiconvergence is ideally suited to study extremum estimators of partially identified models.

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), Fan and Park (2007), Guggenberger, Hahn, and Kim (2008), Rosen (2008), Andrews and Guggenberger (2009), Bugni (2009), Galichon and Henry (2009), Hahn and Ridder (2009), Moon and Schorfheide (2009), Yildz (2009), Andrews and Soares (2010), Canay (2010), and Pakes, Porter, Ho, and Ishii (2011). 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.

We evaluate the finite sample performance of our inference methods through Monte Carlo experiments. For this, we apply our methods to a special case of models studied in Blundell, Kristensen, and Matzkin (2011).

The paper is organized as follows. In section 2, we summarize CHT’s econometric framework and introduce some useful background. We establish the asymptotic distribution of the normalized support function and develop our inference methods in section 3. Section 4 studies moment inequality models 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.


\[^2\]To the best of our knowledge, CHT 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.
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 $S \subset \mathbb{R}^d$, $d \in \mathbb{N}$. The function $f : S \rightarrow \bar{\mathbb{R}}_+$ is proper on $S$ if $f(x) < \infty$ for at least one $x \in S$. If $f$ is proper on $S = \mathbb{R}^d$, we say $f$ is proper.

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

The set $\Theta$ is the parameter space, which we take here to be of finite dimension. Compactness is a standard assumption on $\Theta$ for extremum estimation. Convexity and nonempty interior help us to avoid the “parameters on the boundary problem” for partially identified models.

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 $\Theta$ for all $\omega \in F \in \mathcal{F}$, $P(F) = 1$, and for all $\omega \in \Omega$ and $\theta / \in \Theta$, $Q_n(\omega, \theta) = \infty$.

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 $\{X_i : \Omega \rightarrow \mathbb{R}\}$ is a sequence of random variables and $q$ is a suitable function, e.g., $q(x, \theta) = (x - \theta)^2$ for scalar $x$ and $\theta$. 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 $Q_n(\cdot, \theta)$.

Another common choice for $Q_n$ is that associated with moment inequalities,

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

where $m(x, \theta)$ is a suitable vector-valued function such that $E[m(X_i, \theta)] \leq 0$ for one or more

\footnote{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.}
values of $\theta$, $[z]_+$ truncates each negative component of vector $z$ to 0, 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 $\Theta_I$ satisfies

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

There are numerous examples where the identified set can be written as in (2.2). See Manski and Tamer (2002), Bajari, Benkard, and Levin (2007), Romano and Shaikh (2008, 2010), Ciliberto and Tamer (2009), and Kaido and White (2009). Leading examples are the cases where $\Theta_I$ is a closed interval in $\mathbb{R}$ or an ellipsoid in $\mathbb{R}^2$. $\Theta_I$ is a primary object of interest here. In particular, we are concerned with estimation and inference for $\Theta_I$.

We ensure next that $\Theta_I$ is a compact convex set contained in the interior $\Theta^o$ of $\Theta$.

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

The compactness of $\Theta$ and Assumption 2.2 (i) imply the compactness of $\Theta_I$. 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. Although the convexity of $\Theta_I$ limits the scope of examples we can study to some degree, there are, however, still many examples whose identified set is convex and can be characterized as the set of minimizers of some criterion function. We here present some of them based on simplifications of well known models.

**Example 2.1 (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 $E(X_1) \leq \theta \leq E(X_2)$.

\footnote{For example, entry game models (see for example Tamer, 2003; Ciliberto and Tamer, 2009) typically do not give convex identified sets under commonly used parametric specifications of payoff functions and the distribution of the unobserved heterogeneity.}
Example 2.2 (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 outcome interval \([Y_{1i}, Y_{2i}]\) is observed for each \( i = 1, \cdots, n \). The outcome interval satisfies the following moment inequalities

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

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

\[
Q(\theta) := \int (E(Y_{1i} | X_i = x) - x' \theta)_+^2 + (x' \theta - E(Y_{2i} | X_i = x))^2 dP_0(x). \tag{2.3}
\]

Example 2.3 (Reveal preference bounds): Consider a population of heterogeneous consumers. We represent unobserved heterogeneity by the scalar random variable \( \epsilon : \Omega \rightarrow [0, 1] \). For each income level \( x \in \mathcal{X} \subseteq \mathbb{R}_+ \), price vector \( \pi \in \mathbb{R}^d \), and \( \epsilon \in [0, 1] \), the consumer demand \( y \in \mathbb{R}^d_+ \) is given by:

\[
y = D(x, \pi, \epsilon).
\]

Let \( \{\pi_1, \cdots, \pi_J\} \) be a series of (non-stochastic) observed prices of \( d \) goods. Suppose that for each individual \( i \) and period \( j \), we observe her income \( X_{ij} \) and demand \( Y_{ij} \). As Blundell, Kristensen, and Matzkin (2011) show, the demand function \( \pi \mapsto D(x, \pi, \tau) \) is then identified only on the set \( \{\pi_1, \cdots, \pi_J\} \) of observed prices. When a consumer faces a new price \( \pi_0 \notin \{\pi_1, \cdots, \pi_J\} \), her demand is not point identified but is only set identified. Let \( x_0 \) be given. Any demand response \( \theta \in \Theta := \{\theta \in \mathbb{R}^d_+ : \pi'_0 \theta = x_0\} \) that is consistent with the revealed preference and observed expansion paths then satisfies:

\[
\pi'_j \theta \geq \gamma_{\tau,j}, \quad j = 1, \cdots, J, \tag{2.4}
\]

where for each \( \tau \in [0, 1] \) and \( j \), the intersection income \( \gamma_{\tau,j} \) is defined as the solution to

\[
\pi'_0 D(\gamma, \pi_j, \tau) = x_0. \tag{2.5}
\]

The inequalities (2.4) define the identified set \( \Theta_I \) of demand responses under \( \pi_0 \). Under mild assumptions, \( \Theta_I \) is a non-empty convex set, which can be equivalently defined as the set of
minimizers of the following criterion function:

\[ Q(\theta) := \sum_{j=1}^{J} W_j (\gamma \tau, j - \pi'_j \theta)_+, \quad (2.6) \]

where \( W_j \geq 0, j = 1, \ldots, J \) are suitable weights on the inequalities in (2.4).

**Example 2.4 (Discrete choice):** Suppose an agent chooses \( Z_i \in \mathbb{R}^k \) from a set \( Z := \{z_1, \ldots, z_J\} \) in order to maximize his/her expected payoff \( E[\pi(Y_i, z, \theta_0) | I] \), where \( Y_i \) is a vector of observable random variables, and \( I \) is the agent’s information set. A common specification for the payoff function is \( \pi(y, z, \theta) = \psi(y, z) + z' \theta + \epsilon \) for some known function \( \psi \) and an unobservable error \( \epsilon \). The optimality condition for the agent’s choice is then given by the moment inequalities:

\[
E[\{\psi(Y_i, z_j) - \psi(Y, z_k)\} - (z_j - z_k)' \theta]1\{Z_i = z_k\} \leq 0, \quad j = 1, \ldots, J. \quad (2.7)
\]

The identified set consists of parameter values that satisfy this optimality condition. A population criterion function can then be defined as in (2.1) with \( X_i = (Y_i, Z'_i)' \) and \( m_j(x, \theta) = \{\psi(y, z_j) - \psi(y, z_k)\} - (z_j - z_k)' \theta\}1\{y = z_k\} \).

**Example 2.5 (Pricing kernel):** Let \( Z_i : \Omega \rightarrow \mathbb{R}^J \) be the payoffs of \( J \) securities that are traded at a price of \( V_i \in \mathbb{R}^J_+ \). If short sales are not allowed for any securities, then the feasible set of portfolio weights is restricted to \( \mathbb{R}^J_+ \) and the standard Euler equation does not hold. Instead, under power utility, the following Euler inequalities hold (See Luttmer, 1996):

\[
E[\frac{1}{1 + \rho} Y_i^{-\gamma} Z_i - V_i] \leq 0, \quad (2.8)
\]

where \( Y_i : \Omega \rightarrow \mathbb{R}_+ \) is a state variable, e.g. consumption growth, \( \rho \) is the investor’s subjective discount rate, and \( \gamma \) is the relative risk aversion coefficient. When the payoff \( Z_i \) takes nonnegative values almost surely, the set of parameter values \( \theta = (\rho, \gamma) \) that satisfy (2.8) is convex. A criterion function can then be defined as in (2.1) with \( X_i = (Y_i, Z'_i, V'_i)' \) and \( m_j(x, \theta) = \frac{1}{1 + \rho} y^{-\gamma} z_j - v_j \).

**Remark 2.1:** Examples 2.1, 2.4, and 2.5 belong to the class of moment inequality models studied in Section 4. Example 2.2 can be also reformulated as a moment inequality model when \( X_i \) has a finite support. Even if \( X_i \) does not have a finite support, the framework of this paper is applicable, as the identified set can be characterized as the set of minimizers of a criterion function. Similar to Examples 2.1, 2.4, and 2.5, inequality constraints shape the identified set in Example 2.3. However, the constraints are not imposed on the moments. For this example, our approach is particularly well-suited because a criterion function can be easily constructed, although it appears hard to directly apply existing inference methods for
moment inequality models to this example.

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

\[
\zeta_n(\theta) := a_n Q_n(\theta), \quad \theta \in \mathbb{R}^d.
\]

The constants \( a_n \) normalize the criterion function so that \( \zeta_n \) 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 \( \zeta_n \):

**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 \( \Theta_I \), 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

\[
\tilde{d}_H(A, B) := \sup_{a \in A} d(a, B),
\]

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

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

The directed Hausdorff distance takes the value 0 when \( A \subseteq B \) and a positive value otherwise. 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 \( \Theta_I \) 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.

---

5The directed Hausdorff distance is also called the lower Hausdorff hemimetric. A hemimetric \( d \) defined on a set \( E \) is a mapping \( E \times E \to \mathbb{R} \) such that for any \( x, y, z \in E \), (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 \( \tilde{d}_H(B, A) \).
Assumption 2.3: (i) $\sup_{\theta \in \Theta} \{Q(\theta) - Q_n(\theta)\} = o_p(1)$. (ii) $\sup_{\theta \in \Theta} Q_n(\theta) = O_p(1/a_n)$. (iii) There exist positive constants $(\delta, \kappa, \gamma)$ such that for any $\epsilon \in (0, 1)$, there are $(\kappa_\epsilon, n_\epsilon)$ such that for all $n \geq n_\epsilon$

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

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

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 $t_n \geq \sup_{\theta \in \Theta} a_n Q_n(\theta)$ 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 $\kappa_n$ is a slowly diverging sequence, e.g., $\kappa_n = \log \log n$. Theorem B.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 $\Theta_n$ of $\Theta$, which could be data dependent (i.e., Effros-measurable functions on $\Omega$), 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, $d_H(\hat{\Theta}_n(t_n), \Theta_I) = o_p(1)$. Suppose, in addition, Assumption 2.3 (iii) and 2.4 (ii) hold. Then, $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t_n), \Theta_I) = O_p(1)$.

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 $\Theta_I$ has nonempty interior, the set of points satisfying the moment inequalities approximate $\Theta_I$ 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 \( p \). 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 \( t \in \mathbb{R}_+ \):

\[
s(p, \hat{\Theta}_n(t)) := \sup_{\theta \in \Theta} \langle p, \theta \rangle \tag{2.9}
\]

s.t. \( a_n Q_n(\theta) \leq t \).

When \( Q_n \) is convex in a neighborhood of \( \Theta_I \), this is a convex programming problem, which is easily solvable using standard algorithms (See for example Boyd and Vandenberghe, 2004). Often, such algorithms also find a point \( \hat{\theta}_n(p, t) \) in the support set \( H(p, \hat{\Theta}_n(t)) \) as a solution of the problem (2.9). Therefore, one may trace out the boundary of \( \hat{\Theta}_n(t) \) by solving (2.9) for different values of \( p \).

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 \( \mathbb{R}^d \). Hörmander's theorem (Li, Ogura, and Kreinovich, 2002) ensures that the metric space \( (\mathcal{K}_c, d_H) \) can be isometrically embedded into \( C(\mathbb{S}^{d-1}) \), the space of bounded continuous functions on \( \mathbb{S}^{d-1} \) equipped by the usual uniform norm. Thus, for any \( A, B \in \mathcal{K}_c \), it holds that

\[
d_H(A, B) = \sup_{p \in \mathbb{S}^{d-1}} |s(p, A) - s(p, B)|.
\]

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

\[ Z_n(p, t) := \frac{1}{\gamma_n} \left( s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I) \right). \]

This process is useful for conducting inference. Hörmander’s theorem ensures that when \( \hat{\Theta}_n(t), \Theta_I \in \mathcal{K}_c \), the distance \( \frac{1}{\gamma_n} d_H(\hat{\Theta}_n(t), \Theta_I) \) equals \( \sup_{p \in S^{d-1}} |Z_n(p, t)| \), a functional of \( Z_n(\cdot, t) \). For the directed Hausdorff distance, BM has shown that for any \( A, B \in \mathcal{K}_c \), \( \tilde{d}_H(A, B) = \sup_{p \in S^{d-1}} \{ s(p, A) - s(p, B) \}_+ \). From these results, together with Assumptions 2.1 and 2.2 (i), we have that for given \( t \in \mathbb{R}_+ \)

\[
\frac{1}{\gamma_n} \tilde{d}_H(\Theta_I, \hat{\Theta}_n(t)) = \sup_{p \in S^{d-1}} \{-Z_n(p, t)\}_+ \\
\frac{1}{\gamma_n} \tilde{d}_H(\hat{\Theta}_n(t), \Theta_I) = \sup_{p \in S^{d-1}} \{Z_n(p, t)\}_+
\]

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

As we show, this distribution is a stochastic process on \( S^{d-1} \). 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 \( \mu \) such that for all \( t \in \mathbb{R}_+ \), \( Z^*(\cdot) := Z(\cdot, t) - \mu(t) \) is a mean zero process on \( S^{d-1} \), where \( Z(\cdot, t) \) is the desired weak limit of \( Z_n(\cdot, t) \).

### 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 \( (E, d) \), so that for \( n = 1, 2, ..., \xi_n : \Omega \times E \to \mathbb{R} \) is jointly measurable, where \( \mathbb{R} := \mathbb{R} \cup \{-\infty, \infty\} \). For simplicity, we often suppress the dependence of \( \xi_n \) on \( \omega \in \Omega \), but this should be understood implicitly. In specific contexts, we also view \( \xi_n \) as a mapping from the sample space \( \Omega \) to a space of functions on \( E \).

The simplest convergence in distribution concept for stochastic processes is weak convergence in finite dimensions. For this, we use the notation \( \overset{d}{\to} \) to denote the usual convergence in distribution (weak convergence) for a vector of finite dimension (as in, e.g., White, 2001, p.65). A sequence of stochastic processes \( \{\xi_n, n \geq 1\} \) on \( E \) is said to weakly converge in finite dimension to a limit \( \xi \), denoted \( \xi_n \overset{f.d.}{\to} \xi \), if for any finite \( m \)-tuple \( (x_1, ..., x_m) \), where \( x_j \in E \) for each \( j = 1, ..., m \),

\[
(\xi_n(x_1), ..., \xi_n(x_m)) \overset{d}{\to} (\xi(x_1), ..., \xi(x_m)).
\]

A stronger notion, the weak convergence in the uniform metric is defined as follows.
Definition 2.6 (Weak convergence in the uniform metric): Let \((E, d)\) be a complete separable metric space. Let \(C_0(l^\infty(E))\) be the space of bounded continuous functions on \(l^\infty(E)\). A sequence of stochastic processes \(\{\xi_n, n \geq 1\}\) on \(E\) is said to weakly converge in the uniform metric to a limit \(\xi\), denoted \(\xi_n \xrightarrow{u.d.} \xi\), if for any \(f \in C_0(l^\infty(E))\)

\[
E[f(\xi_n)] \to E[f(\xi)]
\]

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 \(l^\infty(E)\), where \(l^\infty(E)\) is the space of uniformly bounded functions on \(E\); see, e.g., van der Vaart and Wellner (2000). Here, a main goal is to find \(Z(\cdot, t)\) such that \(Z_n(\cdot, t) \xrightarrow{u.d.} Z(\cdot, t)\) for \(Z_n(\cdot, t)\) 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 \(E\) is said to weakly epiconverge to a limit \(\xi\), denoted \(\xi_n \xrightarrow{e.d.} \xi\), if for any compact subsets\(^7\) \(R_1, \ldots, R_k\) of \(E\) with open interiors \(R_1^o, \ldots, R_k^o\) and any finite \(m\)-tuple of real numbers \(\tau_1, \ldots, \tau_m\),

\[
P\left( \inf_{x \in R_1} \xi(x) > \tau_1, \ldots, \inf_{x \in R_m} \xi(x) > \tau_m \right)
\]

\[
\leq \liminf_{n \to \infty} P\left( \inf_{x \in R_1} \xi_n(x) > \tau_1, \ldots, \inf_{x \in R_m} \xi_n(x) > \tau_m \right) \quad (2.10)
\]

\[
\leq \limsup_{n \to \infty} P\left( \inf_{x \in R_1^o} \xi_n(x) \geq \tau_1, \ldots, \inf_{x \in R_m^o} \xi_n(x) \geq \tau_m \right)
\]

\[
\leq P\left( \inf_{x \in R_1^o} \xi(x) \geq \tau_1, \ldots, \inf_{x \in R_m^o} \xi(x) \geq \tau_m \right). \quad (2.11)
\]

We call the condition given by (2.10) the lower epilimit condition. Similarly, we call that given by (2.11) the upper epilimit condition\(^8\).

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\(^9\). 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).

\(^7\)In this definition, the sets \(R_1, \ldots, R_k\) can instead be taken from a class of relatively compact sets \(V\) such that (i) \(V\) is closed under finite union and intersection; (ii) each compact set \(K\) in \(E\) is representable as the intersection of a decreasing sequence in \(V\); and (iii) each open set \(G\) in \(E\) is representable as the union of an increasing sequence in \(V\). A typical example for such a \(V\) is a class of closed rectangles. See Pflug (1992) for details.

\(^8\)These names are motivated by Proposition 7.29 in Rockafellar and Wets (2005).

\(^9\)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 Phillips (2006).
For our purposes, weak epiconvergence of a version of the criterion function \( \zeta_n \) helps ensure the finite dimensional weak convergence of \( Z_n(\cdot, t) \). The desired results then follow by establishing asymptotic tightness of \( \{Z_n(\cdot, t)\} \).

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 \( Z_n(\cdot, t) \) 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 \( Z_n(\cdot, t) \) converges weakly in the uniform metric to a stochastic process on \( S^{d-1} \) under appropriate regularity conditions on \( \tilde{\zeta}_n \). We then present our inference methods using functionals of \( Z_n(\cdot, t) \).

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 \( f : \mathbb{R}^d \to \overline{\mathbb{R}} \) 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 \( f : \Omega \times \mathbb{R}^d \to \overline{\mathbb{R}} \) is such that \( f(\omega, \cdot) \) is lsc for all \( \omega \in F \in \mathcal{F}, P(F) = 1 \), 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 \( s(p, \hat{\Theta}_n(t)) = s(p, \hat{\Theta}_n(t_n)) \) if \( \hat{\Theta}_n(t) = \emptyset \), where \( t_n := \inf_\Theta a_n Q_n(\theta) \). This convention ensures that \( Z_n(\cdot, t) \in C(S^{d-1}) \) a.s. Note that \( P(\hat{\Theta}_n(t) = \emptyset) \to 0 \) 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 \( Q_n(\cdot) \).

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

\[
s(p, \hat{\Theta}_n(t)) < u \iff \inf_{\theta \in K_{u,p} \cap \Theta} a_n Q_n(\theta) > t,
\]

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\(^{10}\). Our goal is then to relate the normalized support function \(Z(p,t)\) to a localized version of the criterion function.

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

\[
\hat{\zeta}_n(\theta, \lambda) := a_n Q_n(\theta + \lambda/a_n^{1/\gamma}), \quad \theta \in \partial \Theta_I, \, \lambda \in \mathbb{R}^d.
\]

The local parameter \(\lambda\) belongs to a shifted and rescaled space \(a_n^{1/\gamma}(\Theta - \theta) := \{\lambda \in \mathbb{R}^d : \lambda = a_n^{1/\gamma}(\hat{\theta} - \theta), \hat{\theta} \in \Theta\}\). For \(\hat{\zeta}_n\), Lemma 3.1 implies that

\[
\{\omega : Z(p,t) < u \} = \left\{ \omega : \inf_{(\theta,\lambda) \in R_{n,u,p}} \hat{\zeta}_n(\theta, \lambda) > t \right\}, \quad (3.1)
\]

where

\[
R_{n,u,p} := \{(\theta, \lambda) : \lambda \in K_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta), \theta \in H(p, \Theta_I)\}.
\]

To analyze the event in (3.1), it is important to understand the behavior of \(R_{n,u,p}\) as \(n\) increases. For this, let \(\partial \Theta\) denote the boundary of \(\Theta\). We call elements of \(\Theta_I \cap \partial \Theta\) identified parameters on the boundary (of \(\Theta\)). The remaining elements of \(\Theta_I\) are identified parameters in the interior (of \(\Theta\)). How \(R_{n,u,p}\) behaves in the limit depends on whether or not there is an identified parameter on the boundary of \(\Theta\).

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)\(^{11}\). Thus, for any \((u,p)\), we have the PK convergences

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

On the other hand, if there is an identified parameter on the boundary of \(\Theta\), the limit of the sequence of graphs \(\{R_{n,u,p}\}_{n=1}^\infty\) has a form that depends on the structure of \(\Theta\). 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. Lemma B.1 provides conditions ensuring that

---

\(^{10}\)Note that if \(\hat{\Theta}_n(t) = \emptyset\), we take \(s(p, \hat{\Theta}_n(t)) = \sup_{p, \theta \in \Theta} (p, \theta) = -\infty\).

\(^{11}\)For a sequence \(\{C_n\}_{n \in \mathbb{N}}\) of subsets of \(\mathbb{R}^d\), the inner limit is the set \(\lim \inf_{n \to \infty} C_n := \{x : \exists \{x_n\}_{n \in \mathbb{N}}\) such that \(x_n \to x \) and \(x_n \in C_n, \forall n\}\) while the outer limit is the set \(\lim \sup_{n \to \infty} C_n := \{x : \exists \{x_n\}_{n \in \mathbb{N}}\) such that \(x_n \to x \) and \(x_n \in C_n, \forall k\}\). The limit of the sequence exists if inner and outer limit sets are equal: \(\lim_{n \to \infty} C_n = \lim \inf_{n \to \infty} C_n = \lim \sup_{n \to \infty} C_n\). 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, Chapter 4) for details.
$R_{n,u,p}$ 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.

Under regularity conditions, the infimum over $R_{u,p}$ can be further approximated by that over a compact set $\tilde{R}_{u,p}$. Lemma B.2 in the appendix ensures that, to study the (finite-dimensional) asymptotic behavior of $Z_n(\cdot,t)$, it suffices to study the asymptotic behavior of the infima of $\tilde{\zeta}_n$ over compact sets, which can be controlled if $\tilde{\zeta}_n$ weakly epiconverges to a known limiting process $\tilde{\zeta}$. If so, we can seek a process $Z$ such that

$$P(Z(p_1,t) < u_1, ..., Z(p_m,t) < u_m)$$

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

The portmanteau theorem then implies $Z_n(\cdot,t) \overset{f.d.}{\to} Z(\cdot,t)$. The next theorem establishes this; it further gives the asymptotic distributions of the Hausdorff distances.

**Theorem 3.1:** Suppose that Assumptions 2.1, 2.2, 2.3, 2.4, and B.1 (in the appendix) 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) $Z_n(\cdot,t) \overset{f.d.}{\to} Z(\cdot,t)$, where $Z(\cdot,t)$ is a stochastic process on $S^{d-1}$, which has the representation

$$Z(p,t) = \sup_{\theta \in H(p,\Theta_I)} s\left(p, \hat{\Lambda}(t,\theta)\right);$$ (3.2)

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

$$P(Z(p_1,t) < u_1, ..., Z(p_m,t) < u_m)$$

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

(iii) $Z_n(\cdot,t) \overset{u.d.}{\to} Z(\cdot,t)$.

By the continuous mapping theorem, any continuous function of $Z_n(\cdot,t)$ also converges weakly. This result can be used to conduct inference for $\Theta_I$ or its elements.
3.2 Inference

Using the asymptotic distribution results of the previous section, we now study inference for \( \Theta_I \) and points in \( \Theta_I \). To the best of our knowledge, BM is the first article that uses the estimated support function for inference. They study confidence collections and a confidence set for the identified set \( \Theta_I \). We first show that the same type of inference is possible. We further contribute the literature by studying a confidence set for a point \( \theta_0 \) in the identified set using the estimated support function.

For making asymptotically valid inference, we need to consistently estimate critical values of the form:

\[
c_{1-\alpha}(t) := \inf \left\{ x : P \left( \sup_{p \in \Psi_0} \Upsilon(Z(p,t)) \leq x \right) \geq 1 - \alpha \right\},
\]

for \( \Psi_0 \subseteq S^{d-1} \) and a known function \( \Upsilon : \mathbb{R} \to \mathbb{R} \). It is often difficult to compute \( c_{1-\alpha}(t) \) directly, as the required asymptotic distribution differs from case to case. Specifically, the properties of \( Z(\cdot, t) \) depend on the weak epilimit \( \tilde{\zeta} \) and therefore on the functional form of the criterion function. Also, the distribution of \( Z(\cdot, t) \) depends on the characteristics of the true identified set \( \Theta_I \). 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; See Kaido and Santos (2011) for such examples.

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 3.1**: Let Assumption 2.1 hold with \( Q_n(\omega, \theta) = \tilde{Q}_n(X_1(\omega), \cdots, X_n(\omega), \theta) \) where \( \tilde{Q}_n : \prod_{i=1}^n \mathbb{R}^k \times \mathbb{R}^d \to \mathbb{R}_+ \) 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. Given \( \{X_i\} \), a straightforward subsampling algorithm is the following.

**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 \( N_{n,b} = \left( \begin{array}{c} n \\ b \end{array} \right) \) denote the number of subsamples of size \( b \) from a sample of size \( n \). Let \( \{\Psi_n\} \) be a sequence of random closed subsets of \( S^{d-1} \).

**Step 1.** For \( k = 1, \cdots, N_{n,b} \), construct \( \tilde{\Theta}_{n,b,k}(t) \), the set estimator for the \( k \)-th subsample, computed as a \( t \)-level set of the criterion function \( \zeta_{n,b,k}(X_{k_1}, \cdots, X_{k_b}, \theta) = a_k \tilde{Q}_{n,b,k}(X_{k_1}, \cdots, X_{k_b}, \theta) \), with the obvious notation.
Step 2. For $k = 1, \ldots, N_{n,b}$, compute

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

Step 3. Compute the $100 \times (1 - \alpha)\%$ 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

$$\hat{F}_{n,b}(x,t) := N_{n,b}^{-1} \sum_{1 \leq k \leq N_{n,b}} 1 \left\{ \sup_{p \in \Psi_n} \Upsilon(Z_{n,b,k}(t)) \leq x \right\}.$$

For any $t$, let $F(x,t) := P(\sup_{p \in \Psi_0} \Upsilon(Z(p,t)) \leq x)$. The next theorem is a basic result for subsampling statistics based on the normalized support function\textsuperscript{12}.

**Theorem 3.2:** Suppose the conditions of Theorem 3.1 and Assumption 3.1 hold. Suppose further that $\Upsilon$ is Lipschitz continuous, $\Psi_0$ is compact, and that $d_H(\Psi_n, \Psi_0) = o_p(1)$. Let $\hat{F}_{n,b}(\cdot, t)$ and $\hat{c}_{n,b,1-\alpha}(t)$ be computed by Algorithm 3.1. Suppose that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$.

(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 except $x = 0$, then $\sup_{|x| \geq \epsilon} |\hat{F}_{n,b}(x,t) - F(x,t)| \to 0$ in probability;

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

$$\lim_{n \to \infty} P \left( \sup_{p \in \Psi_0} \Upsilon(Z_n(p,t)) \leq \hat{c}_{n,b,1-\alpha}(t) \right) = 1 - \alpha.$$

**Remark 3.1:** The consistency result in Theorem 3.2 (ii) is weaker than the uniform convergence of subsampling CDFs over the real line. It establishes the uniform convergence of the subsampling CDF on compact sets excluding 0. As we see below, our statistics have limiting distributions whose CDFs are discontinuous at $x = 0$. 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.

**Remark 3.2:** An advantage of subsampling is that it provides a general inference method that can be applied to any example that satisfies the conditions of Theorem 3.1 and Assumption 3.1 without any further assumptions. We note, however, that if the example of interest has additional structure, an alternative inference method may be preferable in terms of the accuracy of approximation or computational tractability. For example, the generalized moment selection (GMS) procedure in Andrews and Soares (2010) combined with bootstrap

\textsuperscript{12}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.
allows the researcher to construct a uniformly valid confidence region for points in \( \Theta_I \) without losing much power. The score-based weighted bootstrap applied to the class of models defined by convex moment inequalities in Kaido and Santos (2011) does not require repeated set estimation on bootstrap samples and is therefore computationally more efficient.

We consider two types of confidence regions that can be obtained by inverting the hypothesis tests regarding the identified set and its elements. The first confidence set covers \( \Theta_I \) with a prespecified probability asymptotically. This set can be obtained by properly expanding the boundary of the set estimator and proposed first by BM. We show that under some regularity conditions, the choice of \( t \) does not matter asymptotically for constructing this type of confidence set. The second confidence set covers each point in \( \Theta_I \) asymptotically with a prespecified probability. To the best of our knowledge, the general asymptotic distribution of the test statistic used to construct the second confidence set is a new result.

\subsection{Inference for the Identified Set}

Given a set \( A \subseteq \Theta \) and \( \epsilon > 0 \), let \( A^\epsilon \) denote the \( \epsilon \)-expansion of \( A \) defined by \( A := \{ \theta : d(\theta, A) \leq \epsilon \} \). Let \( \Theta_0 \in \mathcal{K}_c \) be a given compact convex set. We first consider testing

\[ H_0 : \Theta_0 \subseteq \Theta_I \quad \text{vs.} \quad H_1 : \Theta_0 \not\subseteq \Theta_I. \tag{3.4} \]

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 \( T_n \rightarrow(t) := a_n^{1/\gamma} \vec{d}_H(\Theta_0, \hat{\Theta}_n(t)) \), which can be written in the form:

\[ T_n \rightarrow(t) = \sup_{p \in \Psi_0} \Upsilon(Z_{n,b,k}(p,t)), \]

with \( \Upsilon(x) = \{-x\}_+ \) and \( \Psi_0 = \mathbb{S}^{d-1} \).

Let \( \hat{c}_{n,b,1-\alpha} \rightarrow \) be computed by Algorithm 3.1 with \( \Psi_n = \mathbb{S}^{d-1} \) for all \( n \). Results for the rejection probability and the consistency against fixed alternatives now follow as a corollary to Theorem 3.2.

\begin{corollary}
Suppose the conditions of Theorem 3.2 hold. Let \( \Theta_0 \) be a nonempty compact convex subset of \( \Theta^0 \).

(i) If \( \Theta_0 \subseteq \Theta_I \) and \( \alpha \in (0, 0.5) \), then \( \hat{c}_{n,b,1-\alpha} \rightarrow(t) = c_{\gamma/\alpha}(t) + o_p(1) \), and the test has asymptotic rejection probability bounded above by \( \alpha \):

\[ \lim_{n \to \infty} P(T_n \rightarrow(t) > \hat{c}_{n,b,1-\alpha} \rightarrow(t)) \leq \alpha; \]

\end{corollary}

\footnote{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 \( a_n^{1/\gamma} \vec{d}_H(\Theta_n(t), \Theta_0) \) (the scaled upper Hausdorff hemimetric).}
(ii) If $\Theta_0 \not\subseteq \Theta_I$, then the test is consistent:

\[
\lim_{n \to \infty} P \left( T_n^+ (t) > c_{n,b,1-\alpha}^+ (t) \right) = 1.
\]

Let $\hat{c}_{n,b,1-\alpha}^+ (t) := \hat{c}_{n,b,1-\alpha}^+ (t)/a_n^{1/\gamma}$. Our one-sided confidence set is given by the $\hat{c}_{n,b,1-\alpha}^+ (t)$-expansion of the set estimator $\hat{\Theta}_n(t)$14. This is a Wald-type one-sided confidence set that is directly comparable to the QLR-type confidence set studied by CHT.

**Theorem 3.3**: Suppose the conditions of Theorem 3.2 hold. Suppose $\alpha \in (0,0.5)$. Then for $t$ small enough,

\[
\lim_{n \to \infty} P \left( \Theta_I \subseteq \hat{\Theta}_n^+ (t) \right) = 1 - \alpha.
\]

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.

**Remark 3.3**: Using the scaled Hausdorff metric $T_n(t) := d_H(\hat{\Theta}_n(t), \Theta_0)$, one may also test a stronger hypothesis $H_0 : \Theta_0 = \Theta_I$ against $H_1 : \Theta_0 \neq \Theta_I$. This type of test is also discussed in BM. An inversion of this test yields two-sided confidence regions $(C_n^I, C_n^O)$ such that $\lim inf_{n \to \infty} P(C_n^I \subseteq \Theta_I \subseteq C_n^O) = 1 - \alpha$. See also Kaido and Santos (2011).

### 3.2.2 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 $Z(p,t)$. In this case, we can re-center the process $Z_n(p,t)$ 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 one-sided confidence sets for $\Theta_I$.

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

\[
t_{1-\alpha}^* := \inf \left\{ t \in \mathbb{R}_+ : c_{1-\alpha}^+ (t) = 0 \right\}.
\]

We will show that, for any $0 \leq t < t_{1-\alpha}^*$, confidence sets constructed in the manner of Theorem 3.3 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 $t < t_{1-\alpha}^*$.

---

14The idea of the test inversion to yield a collection of compact convex sets, called confidence collection, is discussed in detail in BM. It can be shown that $\hat{\Theta}_n^\pm$ can be obtained as the union of elements in the confidence collection.
We start with the following lemma that shows $c_{1-\alpha}^\gamma$ is non-increasing on $[0,t_{1-\alpha}]$.

**Lemma 3.2:** Suppose the conditions of Theorem 3.3 are satisfied. Then, for any $0 \leq t < t' \leq t_{1-\alpha}^*$,

$$0 = c_{1-\alpha}^\gamma(t_{1-\alpha}^*) \leq c_{1-\alpha}^\gamma(t') \leq c_{1-\alpha}^\gamma(t) \leq c_{1-\alpha}^\gamma(0).$$

Recall that a confidence set $\hat{\Theta}_n(t)$ is an expansion of the level set $\hat{\Theta}_n(t)$ by the amount $\hat{c}_{n,b,1-\alpha}(t) = \hat{c}_{n,b,1-\alpha}(t)/a_n^{\gamma}$. Lemma 3.2 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 $Z(p,t) = \mu(t) + Z^*(p)$, 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 $t < t_{1-\alpha}^*$.

**Theorem 3.4:** Suppose the conditions of Theorem 3.3 hold. Suppose that the limiting process takes the form $Z(p,t) = \mu(t) + Z^*(p)$ for each $(p,t) \in S^{d-1} \times \mathbb{R}_+$ where $\mu : \mathbb{R}_+ \to \mathbb{R}$ is an unknown function and that $Z_n(p,t) - Z_n(p,t') = \mu(t) - \mu(t') + o_p(1)$ uniformly in $p$. Then (i) for each $\alpha \in (0,1)$ and $0 \leq t < t_{1-\alpha}^*$,

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

(ii) for each $\alpha \in (0,1)$ and for any $0 \leq t \leq t_{1-\alpha}^*$,

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

Theorem 3.4 raises an interesting research question. CHT construct a confidence set $\hat{\Theta}_n(\hat{c}_{n,b,1-\alpha})$ such that $\lim_{n \to \infty} P(\Theta_i \subseteq \hat{\Theta}_n(\hat{c}_{n,b,1-\alpha})) = 1 - \alpha$, where $\hat{\Theta}^\gamma_{n,b,1-\alpha}$ is a subsampling estimate of the $1 - \alpha$ quantile $\tau_{1-\alpha}^*$ of the limiting distribution of their QLR-statistic $\sup_{\Theta_j} a_n Q_n(\Theta)$. If $t_{1-\alpha}^* = \tau_{1-\alpha}^*$ holds, the confidence sets based on the QLR-approach and our 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 3.2:** (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 $c_{1-\alpha}^\gamma(t_l)$

---

15The 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.
of the scaled directed Hausdorff distance \( a_n^{1/\gamma} \bar{d}_H(\Theta_I, \hat{\Theta}_n(t_1)) \) by Algorithm 3.1 with
\( \Upsilon(x) = \{ -x \} _+ \) and \( \Psi_0 = S^{d-1} \), obtaining \( \hat{c}_{\rightarrow n,b,1-\alpha}(t_1) \). Using \( \hat{c}_{\rightarrow n,b,1-\alpha}(t_1) \), expand \( \hat{\Theta}_n(t_1) \) by \( \hat{e}_{\rightarrow n,b,1-\alpha}(t_1) = \hat{c}_{\rightarrow n,b,1-\alpha}(t_1)/a_n^{1/\gamma} \) to obtain \( \hat{\Theta}_{\rightarrow n,b,1-\alpha}(t_1) \).

**Step 2.** Update the level by setting \( t_{l+1} := \sup_{\theta \in \hat{\Theta}_{\rightarrow n,b,1-\alpha}(t_1)} a_n Q_n(\theta) \).

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

The iterative algorithm can be proved to yield an increasing sequence \( \{ t_l, l = 1, 2, \ldots \} \) that tends to \( t^\ast_{1-\alpha} \). As Theorem 3.4 shows, if the limiting process takes the form \( Z(p, t) = \mu(t) + Z^\ast(p) \), one can stop at Step 1, as the iteration does not provide any first-order asymptotic improvement, although it may provide higher order refinements.

### 3.2.3 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\(^\text{16}\). As the true parameter value cannot be distinguished from any other element of \( \Theta_I \), a relevant question would be whether or not a given parameter value \( \theta_0 \) 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 \( \Theta_I \) with at least a prespecified probability.

Let \( \theta_0 \in \Theta \), and consider testing

\[ H_0 : \theta_0 \in \Theta_I \quad \text{vs.} \quad H_1 : \theta_0 \notin \Theta_I. \quad (3.5) \]

Suppose for the moment that \( \theta_0 \in \partial \Theta_I \) and let \( \Psi_0 \subseteq S^{d-1} \) be the set of maximizers of \( \langle p, \theta_0 \rangle - s(p, \Theta_I) \). \( \Psi_0 \) is not a singleton if \( \theta_0 \) is a kink point. We use the directed Hausdorff distance statistic to test the hypothesis. Given \( \theta_0 \), define the statistic

\[ T_{n, \theta_0}(t) := a_n^{1/\gamma} \bar{d}_H(\{ \theta_0 \}, \hat{\Theta}_n(t)) = \sup_{p \in S^{d-1}} a_n^{1/\gamma} \left\{ \langle p, \theta_0 \rangle - s(p, \hat{\Theta}_n(t)) \right\}_+^+. \quad (3.6) \]

The following theorem characterizes the asymptotic distribution of this statistic.

**Theorem 3.5:** Suppose the conditions of Theorem 3.1 hold. Suppose further that \( \theta_0 \in \partial \Theta_I \). Then,

\[ T_{n, \theta_0}(t) \overset{d}{\to} \sup_{p \in \Psi_0} \{-Z(p, t)\}_+. \quad (3.7) \]

\(^{16}\)See also Woutersen (2006), Fan and Park (2007), and Stoye (2009) for extensions of Imbens and Manski’s (2004) analysis.
Remark 3.4: A statistic closely related to $T_{n,\theta_0}(t)$ is studied in Bontemps, Magnac, and Maurin (forthcoming, Proposition 10). To derive the asymptotic distribution of their statistic, these authors construct a sequence $p_n$ of unit vectors that converges to some $p_0 \in \Psi_0$. Our contribution here is to derive the asymptotic distribution of $T_{n,\theta_0}$ without such a sequence. For this, we note that our statistic can be written as

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

where for any $x : \mathbb{S}^{d-1} \to \mathbb{R}$, $\phi_{\theta_0}(x) := \sup_{p \in \mathbb{S}^{d-1}} \langle p, \theta_0 \rangle - x(p)$. In the appendix, we show that $\phi_{\theta_0}$ belongs to a class of Hadamard directionally differentiable functionals. Theorem 3.1 and a functional $\delta$-method in Shapiro (1991) then establish Theorem 3.5.

The asymptotic distribution of the statistic depends on $\theta_0$ through its dependence on $\Psi_0$. For each $\alpha \in (0, 1)$ and $t \in \mathbb{R}_+$, let $c_{1-\alpha}(\theta_0, t)$ be the $1 - \alpha$ quantile of $\sup_{p \in \Psi_0} \{-Z(p, t)\}_+$. Similar to the inference for $\Theta_I$, we estimate $c_{1-\alpha}(\theta_0, t)$ by subsampling. Aspects specific to pointwise inference is that $\Psi_0$ needs to be estimated from data. Note that $\Psi_0$ is the set of maximizers of a certain criterion function. This allows us to use a level-set estimator to estimate $\Psi_0$. For this, let $\{\kappa_n\}$ be a sequence of positive constants such that $\kappa_n \to \infty$ and $\kappa_n/a_n^{1/\gamma} \to 0$. Define

$$\hat{\Psi}_n := \{p \in \mathbb{S}^{d-1} : \langle p, \theta_0 \rangle = s(p, \hat{\Theta}_n(t)) \leq \sup_{p' \in \Psi_0} \langle p', \theta_0 \rangle - s(p', \hat{\Theta}_n(t))) - \kappa_n/a_n^{1/\gamma}\}.$$  (3.8)

Let $\hat{c}_{n,b,1-\alpha}(\theta_0, t)$ be computed by Algorithm 3.1 with $\Upsilon(x) = \{-x\}_+$ and $\Psi_n = \hat{\Psi}_n$ for all $n$. The following theorem establishes that our test has asymptotic level $\alpha$ and is consistent against any fixed alternative hypothesis.

Corollary 3.2: Suppose the conditions of Theorem 3.2 hold.

(i) If $\theta_0 \in \Theta_I$ and $\alpha \in (0, 0.5)$, then $\hat{c}_{n,b,1-\alpha}(\theta_0, t) = c_{1-\alpha}(\theta_0, t) + o_p(1)$, and the test has asymptotic rejection probability $\alpha$:

$$\limsup_{n \to \infty} P\left(T_{n,\theta_0}(t) > \hat{c}_{n,b,1-\alpha}(\theta_0, t)\right) \leq \alpha.$$

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

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

The confidence set for $\theta_0$ is obtained by inverting the test. Define

$$\hat{\Theta}_{n,b,1-\alpha}(t) := \{\theta_0 \in \Theta : T_{n,\theta_0}(t) \leq \hat{c}_{n,b,1-\alpha}(\theta_0, t)\}.$$  

The following theorem shows that this confidence set has the correct coverage probability.
Theorem 3.6: Suppose the conditions of Theorem 3.2 hold. Then for a given \( \alpha \in (0, 0.5) \),

\[
\inf_{\theta_0 \in \Theta_I} \liminf_{n \to \infty} P \left( \theta_0 \in \tilde{\Theta}_{n,b,1-\alpha}(t) \right) \geq 1 - \alpha.
\]

Note the difference between this confidence set and that for the identified set. To construct \( \tilde{\Theta}_{n,b,1-\alpha}(t) \), we use \( \hat{c}_{n,b,1-\alpha}(\theta_0,t) \) as a critical value, instead of \( \hat{c}_{n,b,1-\alpha}(t) \). Intuitively, the former takes into account how precisely each boundary point of \( \Theta_I \) is estimated. On the other hand, the latter takes into account how precisely the whole boundary of \( \Theta_I \) is estimated.

Remark 3.5: A simple extension of pointwise inference yields a conservative test for a hypothesis that \( \Theta_I \) has a nonempty intersection with a known set \( \Theta_0 \). One may use the test statistic \( \inf_{\theta_0 \in \Theta_0} T_{\theta_0}(t) \) for this purpose. This type of hypothesis testing has been studied in Romano and Shaikh (2008) for parametric models and Santos (2012) for nonparametric models. One may envisage other extensions of pointwise inference. One such possibility is inference on linear functionals of \( \theta_0 \). This extension is straightforward in our framework, as any linear functional of \( \theta_0 \) 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\(^\text{17}\). Examples 2.1, 2.4, and 2.5 fall in this class. 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 \( m_j : \mathbb{R}^k \times \mathbb{R}^d \to \mathbb{R}, j = 1, \ldots, J \), that define the following moment inequality

restrictions.

\[ E(m_j(X; \theta)) \leq 0, \quad j = 1, \ldots, J. \]

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

Let \( m_{\theta} \) 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:

\[
Q(\theta) = \varphi(E(m_{\theta}), W(\theta))
\]

\[
Q_n(\theta) = \varphi(\hat{E}_n(m_{\theta}), \hat{W}_n(\theta)),
\]

where \( \varphi : \mathbb{R}^J \times \bar{\mathcal{P}}_J \to \mathbb{R}_+ \) is a non-negative continuous function of the moment condition and the weighting matrix. For example, CHT and Romano and Shaikh (2008, 2010) 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 \( \mathbb{R}_+^J = \{ x \in \mathbb{R}^J : x_j \leq 0, j = 1, \ldots, J \} \). 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” \( \varphi \).

**Assumption 4.1:** (i) Let \( J \in \mathbb{N} \). \( \varphi : \mathbb{R}^J \times \bar{\mathcal{P}}_J \to \mathbb{R}_+ \) 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 \( j = 1, \ldots, J \), and \( \varphi(y, w) = \infty \) if \( y \) or \( w \) contains an infinite element. (ii) Let \( \Theta \subset \mathbb{R}^d, d \in \mathbb{N} \), be compact and convex with nonempty interior. (iii) Let \( W : \mathbb{R}^d \to \bar{\mathcal{P}}_J \) be a measurable mapping, and suppose that \( W \) is finite and continuous on \( \Theta \) and that if \( \theta \notin \Theta \) then \( \det(W(\theta)) = \infty \). (iv) Let \( k \in \mathbb{N} \); for each \( j = 1, \ldots, J \), \( m_j : \mathbb{R}^k \times \mathbb{R}^d \to \mathbb{R} \) is jointly measurable, and for each

\(^{18}\)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 (2010).
\( x \in \mathbb{R}^k \), if \( \theta \notin \Theta \) then \( m_j(x, \theta) = \infty \). (v) Let \((\Omega, \mathcal{F}, P)\) be a complete probability space. Let \( \{X_i : \Omega \rightarrow \mathbb{R}^k\} \) be a sequence of identically distributed random vectors such that for each \( \theta \in \Theta \) and \( j = 1, \cdots, J \), \( E(m_j(X_i, \theta)) < \infty \). (vi) Let \( \tilde{W}_n : \Omega \times \mathbb{R}^d \rightarrow \mathcal{P}_J \) be jointly measurable, and suppose that for each \( \omega \in \Omega \), \( \tilde{W}_n(\omega, \cdot) \) is finite and continuous on \( \Theta \), uniformly in \( n \), and for each \( \omega \in \Omega \), if \( \theta \notin \Theta \) then \( \det \tilde{W}_n(\omega, \theta) = \infty \). (vii) Define \( Q(\theta) := \varphi(E(m_\theta), W(\theta)) \) and \( Q_n(\theta) := \varphi(\tilde{E}_n(m_\theta), \tilde{W}_n(\theta)) \).

Assumption 4.1 ensures that Assumption 2.1 holds for moment inequality models. The assumed continuity of \( W \) on \( \Theta \) and its behavior outside of \( \Theta \) 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 \( \varphi \) is a nonnegative function and that \( E(m_\theta) \) and \( W(\theta) \) are finite on \( \Theta \). 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 \( \Theta \); (ii) \( \{ \theta \in \Theta : \varphi(E(m_\theta), W(\theta)) = 0 \} \subset \Theta^0 \).

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 \( \Theta \).

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 \( \varphi \). 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 \( \epsilon \)-contraction of \( \Theta_I \), 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 \( h_1 : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) such that for any \( w \in \mathcal{P}_J \) and \( x, x^* \in \mathbb{R}^J \), \( |\varphi(x, w) - \varphi(x^*, w)| \leq L_1 h_1(||x - x^*||) \), and there exist \( 0 < L_2 < \infty \) and a continuous increasing function \( h_2 : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) such that for any \( x \in \mathbb{R}^J \) and \( w, w^* \in \mathcal{P}_J \), \( |\varphi(x, w) - \varphi(x, w^*)| \leq L_2 h_2(\max_{ij}(w_{ij} - w_{ij}^*)) \); (i-b) \( \{ m_\theta : \theta \in \Theta \} \) is a P-Donsker class, and \( \tilde{W}_n(\theta) - W(\theta) = o_p(1) \) uniformly over \( \Theta \); (ii-a) \( \sup_{\Theta_I} Q_n(\theta) = O_p(1/\alpha_n) \); (ii-b) There exist positive constants \( (C_1, \delta) \) such that for any \( \theta \in \Theta \), \( \|E(m_{\theta})\|_2 \geq C_1 \min\{d(\theta, \Theta_I), \delta\} \); (ii-c) There exist positive constants \( (C_2, \gamma) \) such that for any \( \omega \in \mathcal{P}_J \) and \( x \in \mathbb{R}^J \), \( \varphi(x, w) \geq C_2 ||x||_\gamma^\gamma \); (ii-d) There exist positive constants \( (C_3, C_4, \epsilon) \) such that for any \( 0 < \epsilon \leq \tilde{\epsilon} \) and \( \theta \in \Theta_I^\epsilon \), \( \max_{1 \leq j \leq J} E(m_{j,\theta}) \leq -C_3 \epsilon \), and \( d_H(\Theta_I^\epsilon, \Theta_I) \leq C_4 \epsilon \).

Assumptions 4.3 (i-a,b) are sufficient for the uniform convergence of \( Q_n \) on \( \Theta \). Assump-

---

\(^{19}\)Strictly speaking, one needs to establish the measurability of \( d_H(\Theta_n(t), \Theta_I) \) to discuss consistency. It is known that the measurability of \( \Theta_n(t) \) as a random closed set is sufficient for this purpose.

\(^{20}\)If we further assume that \( \theta \mapsto \varphi(\tilde{E}_n, \tilde{W}_n(\theta)) \) is globally convex on \( \Theta \), a weaker assumption that \( \tilde{E}_n(\theta) \) and \( W_n(\theta) \) converge in probability pointwise is sufficient as in Andersen and Gill (1982, Corollary II.2) and
tion 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 $\theta$ is outside $\Theta_I$. 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 $\Theta_I^e$, on which the sample criterion function $a_nQ_n(\theta)$ 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 $t \geq \sup_{\Theta} a_nQ_n(\theta)$.

The next step is to show that $\tilde{\zeta}_n(\theta, \lambda) = nQ_n(\theta + \lambda/a_n^{1/\gamma})$ satisfies the local process regularity conditions given in Assumption B.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 $\varphi(x, w) := \|w^{1/2}x\|_+^2 := \|w^{1/2}(x \circ s)\|^2$. 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 $m_{\theta}$ allows a first-order expansion $m_{\theta^*} = m_{\theta} + \nabla m_{\theta}(\theta^* - \theta) + o(|\theta^* - \theta|)$ on $\Theta^o$, where $\nabla m_{\theta}$ is a $d$-by-$J$ matrix and $o(|\theta^* - \theta|)$ represents a small order term. Under these assumptions, we can write

$$\tilde{\zeta}_n(\theta, \lambda) = \left\| \sqrt{n} \hat{E}_n(m_{\theta} + \lambda \sqrt{n}) \hat{W}_n^{1/2}(\theta + \lambda/\sqrt{n}) \right\|_+^2$$

$$= \left\| \sqrt{n} \hat{E}_n(m_{\theta}) + \hat{E}_n(\nabla m_{\theta}) \lambda \hat{W}_n^{1/2}(\theta + \lambda/\sqrt{n}) \right\|_+^2 + o_p(1)$$

$$= \left\| [G_n m_{\theta} + \hat{E}_n(\nabla m_{\theta}) \lambda + \sqrt{n} E m_{\theta}] \hat{W}_n^{1/2}(\theta + \lambda/\sqrt{n}) \right\|_+^2 + o_p(1)$$

$$= \left\| M_n(\theta, \lambda) \hat{W}_n^{1/2}(\theta + \lambda/\sqrt{n}) \right\|_+^2 + o_p(1),$$

where we define $G_n := \hat{\Theta}_n - E$ and $M_n(\theta, \lambda) = G_n m_{\theta} + \hat{E}_n(\nabla m_{\theta}) \lambda + \sqrt{n} E m_{\theta}$.

By the $P$-Donsker property of the moment functions, $G_n m_{\theta} \xrightarrow{u.d.} \mathcal{G}(\theta)$ in $l^\infty(\Theta)$, where $\mathcal{G}$ is a $J \times 1$ zero-mean Gaussian process with almost surely continuous paths, and $\text{Var}(G_j(\theta)) > 0$ 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 $(\theta, \lambda) \in \Theta^o \times \mathbb{R}^d$, $M_n(\theta, \lambda) \xrightarrow{f.d.} \mathcal{M}(\theta, \lambda)$.

Newey and McFadden (1994, Theorem 2.7).
$G(\theta) + \Pi(\theta)\lambda + \zeta(\theta)$ and $\hat{W}_n(\theta + \lambda/\sqrt{n}) \xrightarrow{P} W(\theta)$, where $\zeta(\theta)$’s $j$-th component satisfies

$$
\zeta_j(\theta) = \begin{cases} 
-\infty & \text{if } E(m_j, \theta) < 0 \\
0 & \text{if } E(m_j, \theta) = 0 \\
\infty & \text{if } E(m_j, \theta) > 0
\end{cases}
$$

(4.1)

for $j = 1, 2, \ldots, J$. The components of $\zeta(\theta)$ 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 $\varphi$, 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 $\theta$ and $\lambda$. Then the continuous mapping theorem implies

$$
\tilde{\zeta}_n(\theta, \lambda) \xrightarrow{fd} \varphi(\mathcal{M}(\theta, \lambda), W(\theta))
$$

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 epilimit 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 $j = 1, \ldots, J$, and $x \in \mathbb{R}^k, m_j(x, \cdot)$ is continuously differentiable with respect to $\theta$ on $\Theta^o$ with a continuous gradient $\nabla m_\theta(x, \cdot) \in \mathbb{R}^{d \times J}$, and for some continuous mapping $\Pi : \Theta^o \mapsto \mathbb{R}^{J \times d}$ and each $\theta$ in $\Theta^o$, $\hat{E}_n(\nabla m_\theta) = \Pi(\theta) + o_{as}(1)$ and for each $\alpha \in \mathbb{R}$, $\{\lambda \in \mathbb{R}^d : \Pi(\theta) \lambda \leq \alpha, \theta \in \Theta^o\}$ is bounded; (i-b) $\sqrt{\gamma}E(m_\theta) = \zeta(\theta) + o_p(1)$ for each $\theta$ in $\Theta$, where $\zeta$ is defined by Eq. (4.1); (i-c) $\hat{W}_n(\theta) - W(\theta) = o_{as}(1)$ uniformly over $\Theta$; (ii) The map $\theta \mapsto \varphi(\hat{\mathcal{E}}_n(m_\theta), \hat{W}_n(\theta))$ is convex in a neighborhood of $\Theta^o$; (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 B.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 B.1 are satisfied with weak epilimit $\tilde{\zeta}(\theta, \lambda) :=$

---

\(^{21}\)The mathematical appendix summarizes Knight’s (1999) 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 (1994) for details.
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 \( \sqrt{n}d_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \|Z(\cdot, t)\|_{C(S^{d-1})} \), and \( \sqrt{n}d_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \sup_{p \in S^{d-1}} \{-Z(\cdot, t)\}^+ \), where \( Z(\cdot, t) \) can be represented as
\[
Z(p, t) = \sup_{\theta \in H(p, \Theta_I)} \sup_{\lambda : \varphi(M(\theta, \lambda), W(\theta)) \leq t} \langle p, \lambda \rangle. \tag{4.2}
\]

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

### 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 \( Z(\cdot, t) \) can be derived; (ii) for each \( p \), the limiting process \( Z(p, t) \) 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 \( Z(p, t) = \mu(t) + Z^*(p) \); 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 \( J(\theta) \subseteq \{1, \cdots, J\} \) be the set of indices associated with active moment inequalities, i.e., \( E(m_{j, \theta}) = 0 \) for all \( j \in J(\theta) \). We denote by \( J(\theta) \) the number of elements in \( J(\theta) \). Similarly, let \( J^c(\theta) \subseteq \{1, \cdots, J\} \) collect indices associated with slack moment inequalities at \( \theta \in \partial \Theta_I \), i.e., \( E(m_{j, \theta}) < 0 \) for all \( J^c(\theta) \).

Let \( \Pi_{J(\theta)}(\theta) \) denote the \( J(\theta) \times d \) matrix that stacks rows of \( \Pi(\theta) \) whose indices belong to \( J(\theta) \). Similarly, let \( G_{J(\theta)} \) denote the \( J(\theta) \times 1 \) vector of Gaussian processes that stacks components of \( G \) whose indices belong to \( J(\theta) \). Let \( W_{J(\theta)} \) denote the \( J(\theta) \times J(\theta) \) matrix that collects \( (i, j) \) elements of \( W(\theta) \) for \( i, j \in J(\theta) \).

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

$$\sup_{\lambda} \langle p, \lambda \rangle \quad (4.3)$$

$$s.t. \quad \|W_{\mathcal{J}(\theta)}^{1/2}(\theta)[G_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda]\|^2_+ \leq t.$$ 

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

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

**Assumption 4.5:** (i) For each $\theta \in \partial \Theta_I$, $\text{rank}(\Pi_{\mathcal{J}(\theta)}) = J(\theta)$, i.e. the rows of the Jacobian matrices are linearly independent; (ii) For each $\theta \in \partial \Theta_I$ and $p \in S^{d-1}$, there exists a vector $\eta \in \mathbb{R}_+^{J(\theta)} \setminus \{0\}$ such that $p = \Pi_{\mathcal{J}(\theta)}^\prime \eta$; (iii) For each $\theta \in \partial \Theta_I$, $J(\theta) \leq d$; (iv) Suppose $\varphi(x, w) = \|w^{1/2}x\|^2_+.$

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:

$$\sup_{\theta} \quad \langle p, \theta \rangle \quad (4.4)$$

$$s.t. \quad E(m_{j, \theta}) \leq 0, \text{ for } j = 1, \cdots, J.$$ 

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. Then the process $Z(\cdot, t)$ in Corollary 4.1 can be represented as

$$Z(p, t) = \sup_{\theta \in H(p, \Theta_I)} \left\{ \|R(p, \theta)\|t^{1/2} - \langle R(p, \theta), W_{\mathcal{J}(\theta)}^{1/2}(\theta)G_{\mathcal{J}(\theta)} \rangle \right\}, \quad (4.5)$$

where

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

Furthermore, suppose $W(\theta)$ satisfies $W_{\mathcal{J}(\theta)}(\theta) = [\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)^{-1}]^{-1}$ for any $\theta \in \partial \Theta_I$. Then the limiting process takes the form $Z(p, t) = \mu(t) + Z^*(p)$ with $\mu(t) = t^{1/2}$ and $Z^*(p) = \sup_{\theta \in H(p, \Theta_I)} -[\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)^{-1}]^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p, G_{\mathcal{J}(\theta)}(\theta)).$
Equation (4.5) shows the limiting process $Z(\cdot, t)$ depends on the multivariate Gaussian process $G$, but again we note that the only selected components of $G$ are relevant. Therefore, for each $p \in S^{d-1}$, 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, 2010).

If the weighting matrix satisfies $W_{J(\theta)}(\theta) := [I_{J(\theta)}(\theta)I_{J(\theta)}(\theta)^T]^{-1}$ at each boundary point, then straightforward algebra shows $\|R(p, \theta)\| = 1$, which makes the first term in Eq. (4.5) independent of $\theta$. With this choice of weighting matrix, the limiting process takes the form $Z(p, t) := t^{1/2} + Z^*(p)$.

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: $\sup_{\Theta_I} a_n Q_n(\theta)$. The QLR statistic can be written as

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

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

$$L_n(p, u) := \sup_{\theta \in H(p, \Theta_I)} \sup_{\lambda \in K_{u,p}^-} \tilde{\zeta}_n(\theta, \lambda),$$

where $K_{u,p}^- := \{ \lambda \in \mathbb{R}^d : \langle p, \lambda \rangle \leq u \}$. Note that $\sup_{p \in S^{d-1}} L_n(p, 0) = \sup_{\theta \in \partial \Theta_I} a_n Q_n(\theta)$. We therefore study the asymptotic behavior of the process $L_n(\cdot, u)$ to study that of the QLR statistic. The following theorem establishes the weak convergence of $L_n(\cdot, u)$. 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 B.2 holds. Then $L_n(\cdot, u) \stackrel{\mathbb{D}}{\rightarrow} L(\cdot, u)$ for each $u$, and the process $L$ can be represented as

$$L(p, u) = \sup_{\theta \in H(p, \Theta_I)} \|R(p, \theta)\|^{-1} \left( \left\langle R(p, \theta), W_{J(\theta)}(\theta) G_{J(\theta)}(\theta) \right\rangle + u \right)^\frac{1}{2}.$$

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 $t^*_{1-\alpha} := \inf\{t : \sup_{p \in S^{d-1}} L(p, t) \leq 1\}$.\(^{22}\)

\(^{22}\)In sample, one may use a sample analog $\hat{W}_{n, J_n(\theta)}(\theta) := (\hat{E}_{n, J_n(\theta)}[\nabla m_\theta \hat{E}_{n, J_n(\theta)}[\nabla m_\theta]])^{-1}$ to construct $Q_n$. Here, for each $n$, $J_n(\theta)$ is a mapping from $\Theta$ to a subset of $\{1, \cdots , J\}$ that selects (approximately) binding sample moment conditions at $\theta$. Such moment selection mechanisms are studied in Andrews and Soares (2010).
\[ P(\sup_{p \in \mathbb{S}^{d-1}} \{-Z(p, t)\} \leq 0) \geq 1 - \alpha, \] 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 \( W_J(\theta) = [\Pi J(\theta)(\theta)\Pi J(\theta)(\theta)]^{-1} \) for each \( \theta \in \partial \Theta_T \). Suppose \( \Theta_T \) is strictly convex. Then for each \( p \in \mathbb{S}^{d-1} \), let \( \theta_I(p) \in \partial \Theta_T \) be the boundary point of \( \Theta_T \) such that \( H(p, \Theta_I) = \{\theta_I(p)\} \).

Then, (i)
\[
\sup_{p \in \mathbb{S}^{d-1}} \{-Z_n(p, t) + t^{1/2}\} \xrightarrow{d} Z \quad \text{and} \quad \sup_{\theta_I} nQ_n(\theta) \xrightarrow{d} \mathbf{Z},
\]
where
\[
\mathbf{Z} := \sup_{p \in \mathbb{S}^{d-1}} \left[ \left( (\Pi J(\theta_I(p))(\theta_I(p)))\Pi J(\theta_I(p))(\theta_I(p)) \right)^{-1} \Pi J(\theta_I(p))(\theta_I(p))p, C J(\theta_I(p))(\theta_I(p)) \right]^2.
\]

(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, a continuous functional of a Gaussian process. The second result also has important consequences. It implies the asymptotic equivalence of the Wald and QLR confidence sets for \( \Theta_I \). This is due to Theorem 3.4. When \( t_{1-\alpha}^* = \tau_{1-\alpha}^* \), Theorem 3.4 implies
\[
d_H \left( \hat{\Theta}_n^{\alpha,b,1-\alpha}(t), \hat{\Theta}_n(\tau_{1-\alpha}^*) \right) = o_p(n^{-1/2}),
\]
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.

Using Example 2.1, we illustrate our equivalence results in more detail and give a new interpretation to the results established by BM. Let \( \theta_1 = E(X_{1i}) \) and \( \theta_2 = E(X_{2i}) \). The identified set for \( \theta \) is a closed interval \( \Theta_I = [\theta_1, \theta_2] \). \( \Theta_I \) can be characterized as a set of minimizers of the criterion function \( Q(\theta) = \varphi(E(m_\theta), W(\theta)) \), where \( m_\theta = (X_1 - \theta, \theta - X_2)' \) and \( \varphi \) is as in Assumption 4.5 (iv). Define the sample criterion function by \( Q_n(\theta) = \varphi(\hat{E}_n(m_\theta), \hat{W}_n(\theta)) \). 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 4.1-4.5.

The following results follow immediately from Corollaries 4.1 and 4.2.
Corollary 4.4: Let \( t \) be a 2-by-1 vector of ones. Let \( t \in \mathbb{R}_+ \). Suppose
\[
(\sqrt{n} \hat{E}_n(X_{1i}) - \theta_1), \sqrt{n} (\hat{E}_n(X_{2i}) - \theta_2)' \overset{d}{\to} N(0, \Omega)
\]
and \(-\infty < \theta_1 < \theta_2 < \infty\).

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

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 \( Z(\cdot,t) \). 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 4.3: Let the assumptions of Theorem 4.4 hold. Let \( W_n := \sqrt{n}d_H(\Theta_T, \hat{\Theta}_n(0)) \). Let \( QLR_n := \sup_{\Theta \in \Theta} nQ_n(\theta) \) be CHT’s QLR statistic. Let \( \hat{W}_n := \sqrt{n}d_H(\Theta_T, \hat{\Theta}_n) \) be BM’s Wald statistic, where \( \hat{\Theta}_n = n^{-1} \bigoplus_{i=1}^n F_i \) and \( F_i = [X_{1i}, X_{2i}] \) for \( i = 1, \ldots, n \). Let \( Z \) be the process given in Corollary 4.4. Then
\[
W_n^2 \overset{d}{\to} \max\{(-Z(-1,0))^2, (-Z(1,0))^2\} \quad (4.6)
\]
\[
QLR_n \overset{d}{\to} \max\{(-Z(-1,0))^2, (-Z(1,0))^2\} \quad (4.7)
\]
\[
\hat{W}_n^2 \overset{d}{\to} \max\{(-Z(-1,0))^2, (-Z(1,0))^2\}. \quad (4.8)
\]

The asymptotic equivalence of CHT’s QLR statistic and BM’s Wald statistic in equations (4.7) and (4.8) is due to BM’s Theorem 3.1. Here, Theorem 4.3 adds eq. (4.6).

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 \( W_n^2 \) and \( QLR_n \) 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

\(^{23}\)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.
estimator with \( t = 0 \). Thus, the “exact” equivalence of \( W^2_n \) and \( \tilde{W}^2_n \) 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 \( \Theta_I \), 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.

5 Monte Carlo Experiments

We conduct Monte Carlo experiments using Example 2.3 to examine the performance of our inference method. In what follows, for any vector \( w \), we let \( w^{(k)} \) denote its \( k \)-th component. Following Blundell, Kristensen, and Matzkin (2011), we use a random coefficient Cobb-Douglas structure to generate individual demands. Specifically, for each individual \( i \) and period \( j \), we generate individual income \( X_{ij} \) as an IID normal random variable with mean \( \mu_x \) and variance \( \sigma_x^2 \). We then generate \( Y_{ij} \) as

\[
Y_{ij}^{(k)} = \alpha_i^{(k)} \frac{X_{ij}}{\pi_j^{(k)}}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, J, \quad k = 1, \ldots, d - 1, \quad (5.1)
\]

\[
Y_{ij}^{(d)} = \left(1 - \sum_{k=1}^{d-1} \alpha_i^{(k)} \frac{X_{ij}}{\pi_j^{(d)}}\right), \quad i = 1, \ldots, n, \quad j = 1, \ldots, J, \quad (5.2)
\]

where, for each \( k \), \( \alpha_i^{(k)} \) follows an IID normal distribution with mean \( \mu_{\alpha_i^{(k)}} \) and variance \( \sigma_{\alpha_i^{(k)}}^2 \), and \( \alpha_i^{(h)} \) is independent of \( \alpha_i^{(k)} \) for any \( h \neq k \). The demand function \( D \) is then given by

\[
D^{(k)}(x, \pi, \tau) = \Phi^{-1}_{\mu^{(k)}_\alpha, \sigma^{(k)}_\alpha}(\tau) \frac{x}{\pi^{(k)}}, \quad k = 1, \ldots, d - 1, \quad (5.3)
\]

\[
D^{(d)}(x, \pi, \tau) = \left(1 - \sum_{k=1}^{d-1} \Phi^{-1}_{\mu^{(k)}_\alpha, \sigma^{(k)}_\alpha}(\tau) \right) \frac{x}{\pi^{(d)}}, \quad (5.4)
\]

where for each \( \mu \in \mathbb{R} \) and \( \sigma^2 > 0 \), \( \Phi^{-1}_{\mu, \sigma^2}(\tau) \) denotes the \( \tau \)-th quantile a normal distribution with mean \( \mu \) and variance \( \sigma^2 \). Blundell, Kristensen, and Matzkin (2011) estimate the demand function nonparametrically assuming that the form of \( D \) is unknown. To keep a tight focus on the inference for the identified set of demands, we here assume that the functional form of the demand function is known and use a parametric estimator of \( D \). Specifically, our estimator \( \hat{D}_n \) of the demand functions is obtained by replacing \( \mu^{(k)}_\alpha \) and \( \sigma^{(k)}_\alpha \) in (5.3)-(5.4)
with the following estimators:
\[ \hat{\mu}_n^{(k)} := \frac{1}{n} \sum_{i=1}^{n} \frac{X_{i1}}{n_1^{(k)}} \cdot \hat{\nu}_n^{(k)} := \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_{i1}}{n_1^{(k)} - \hat{\mu}_n^{(k)}} \right)^2, \quad k = 1, \ldots, d - 1. \]

Given \( \hat{D}_n \), the intersection incomes \( \{\gamma_{\tau,j}\}_{j=1}^{J} \) can be estimated as the solutions to \( p_j^{(k)} \hat{D}_n(x, \pi_j, \tau) = x_0 \) for \( j = 1, \ldots, J \). In our setting, the estimator \( \hat{\gamma}_{\tau,j,n} \) admits the closed form:
\[
\hat{\gamma}_{\tau,j,n} := x_0 \left( \sum_{k=1}^{d-1} \Phi^{-1}(\hat{\mu}_n^{(k)}, \hat{\nu}_n^{(k)}) (\pi_j) \frac{\pi_j^{(k)}}{\pi_j} + \left( 1 - \sum_{k=1}^{d-1} \Phi^{-1}(\hat{\mu}_n^{(k)}, \hat{\nu}_n^{(k)}) (\tau) \frac{\pi_j^{(k)}}{\pi_j} \right) \right)^{-1}, \quad j = 1, \ldots, J.
\]

It is straightforward to show that \( \sqrt{n}(\hat{\gamma}_n - \gamma_0) \) is asymptotically normal under regularity conditions. Let \( \Theta := \{ \theta \in \mathbb{R}^d_+ : \pi'\theta = x_0 \} \). For each \( \theta \), define the sample criterion function by
\[
Q_n(\theta) := \sum_{j=1}^{J} \hat{W}_{jn}(\hat{\gamma}_{\tau,j,n} - \pi_j^{(d)})_+, \quad \text{if } \theta \in \Theta, \quad \text{and } Q_n(\theta) = \infty, \quad \text{if } \theta \notin \Theta,
\]
where we simply take \( \hat{W}_{jn} = 1 \) for all \( j \) and \( n \). It can be shown that Assumptions 2.1-2.4 hold with \( a_n = \sqrt{n}, \gamma = 1, \) and \( \Theta_n = \{ \theta \in \Theta : d(\theta, \Theta \setminus \Theta_I) \geq \epsilon_n \}, \epsilon_n = O_p(n^{-1/2}) \). Thus, we may set \( t = 0 \) to construct a \( \sqrt{n} \)-consistent estimator of \( \Theta_I \). \( \hat{\Theta}_n(0) \) is then computed by numerically solving (2.9)\(^{24}\).

In order to examine the performance of our procedure under different numbers of constraints and parameters, we consider six different specifications listed in Table 1. The prices of consumption goods are taken from Table A.I in Blundell, Browning, and Crawford (2008), which are constructed from the British Family Expenditure Survey (FES). For Specifications I-III, we use the prices of food and services from 1983 to 1983+\( J - 1 \). For Specifications IV-VI, we use the price of non-durable goods in addition to these two prices\(^{25}\).

Table 2 summarizes the coverage probabilities of our confidence set for \( \Theta_I \) under specifications I-III. For each specification, the coverage probabilities are computed using subsampling critical values with \( b = 20, 50, 100, \) and 200. Overall, the coverage probabilities are close to the nominal level but slightly conservative when the number of constraints is not large \( (J = 5, 10) \). We also note that the coverage probabilities do not vary much across different subsample sizes. When the number of constraints is relatively large \( (J = 15) \), the coverage probabilities are below the nominal level for some cases, but the magnitude of size distortion is limited (at most 0.006).

Table 3 summarizes the results under Specifications IV-VI. The overall results are simi-

\(^{24}\)We employ the open software Matlab toolboxes YALMIP and MPT, available at http://users.isy.liu.se/johanl/yalmip/ and http://control.ee.ethz.ch/~mpt/.

\(^{25}\)See Blundell, Browning, and Crawford (2008) for details on the definition of the consumption goods.
lar, but the magnitude of size distortion is typically larger than the previous case (at most 0.024). In sum, the size control seems more difficult when the number of parameters and inequality constraints are large relative to the sample size. This is partly in accord with fixed $J$ asymptotics providing poor approximations when the number of inequalities is large (Menzel, 2008).

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 tractability of the support function approach.

We consider the general case where the identified set $\Theta_I$ 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 (forthcoming). Specifically, given a compact convex set $\Theta_0$ or a point $\theta_0$, we present tests for set inclusion $H_0: \Theta_0 \subseteq \Theta_I$ and point inclusion $H_0: \theta_0 \in \Theta_I$.

The test for set inclusion can be inverted to construct 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), Andrews and Guggenberger (2009), and Andrews and Soares (2010).

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, 2010) 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.
References


[40]


### A Tables

**Table 1: Monte Carlo Design**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>2</td>
</tr>
<tr>
<td>$J$</td>
<td>5</td>
</tr>
<tr>
<td>$\mu_x$</td>
<td>39.05</td>
</tr>
<tr>
<td>$\sigma_x^2$</td>
<td>1.50</td>
</tr>
<tr>
<td>$\mu_\alpha^{(1)}$</td>
<td>0.30</td>
</tr>
<tr>
<td>$\mu_\alpha^{(2)}$</td>
<td>0.33</td>
</tr>
<tr>
<td>$v_\alpha^{(1)}$</td>
<td>0.05</td>
</tr>
<tr>
<td>$v_\alpha^{(2)}$</td>
<td>0.05</td>
</tr>
<tr>
<td>$\pi_0^{(1)}/\pi_0^{(2)}$</td>
<td>0.85</td>
</tr>
<tr>
<td>$\pi_0^{(1)}/\pi_0^{(3)}$</td>
<td>1.20</td>
</tr>
<tr>
<td>$x_0$</td>
<td>40</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.50</td>
</tr>
</tbody>
</table>
### Table 2: Coverage probabilities of confidence set \((d = 2)\)

<table>
<thead>
<tr>
<th>(n = 1,000)</th>
<th>(J)</th>
<th>(b)</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.954</td>
<td>0.953</td>
<td>0.951</td>
<td>0.949</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.954</td>
<td>0.957</td>
<td>0.954</td>
<td>0.956</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.954</td>
<td>0.954</td>
<td>0.956</td>
<td>0.954</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 500)</td>
<td>5</td>
<td>0.964</td>
<td>0.960</td>
<td>0.962</td>
<td>0.960</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.966</td>
<td>0.960</td>
<td>0.966</td>
<td>0.964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.948</td>
<td>0.948</td>
<td>0.948</td>
<td>0.950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 250)</td>
<td>5</td>
<td>0.964</td>
<td>0.964</td>
<td>0.960</td>
<td>0.956</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.964</td>
<td>0.964</td>
<td>0.964</td>
<td>0.964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.944</td>
<td>0.952</td>
<td>0.952</td>
<td>0.952</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Coverage probabilities are computed based on 1,000 simulation replications.

### Table 3: Coverage probabilities of confidence set \((d = 3)\)

<table>
<thead>
<tr>
<th>(n = 1,000)</th>
<th>(J)</th>
<th>(b)</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.953</td>
<td>0.954</td>
<td>0.954</td>
<td>0.953</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.952</td>
<td>0.953</td>
<td>0.950</td>
<td>0.952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.935</td>
<td>0.935</td>
<td>0.935</td>
<td>0.934</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 500)</td>
<td>5</td>
<td>0.956</td>
<td>0.958</td>
<td>0.960</td>
<td>0.960</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.944</td>
<td>0.946</td>
<td>0.946</td>
<td>0.942</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.926</td>
<td>0.930</td>
<td>0.928</td>
<td>0.930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 250)</td>
<td>5</td>
<td>0.960</td>
<td>0.956</td>
<td>0.960</td>
<td>0.964</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.956</td>
<td>0.948</td>
<td>0.948</td>
<td>0.956</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.940</td>
<td>0.932</td>
<td>0.932</td>
<td>0.936</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Coverage probabilities are computed based on 1,000 simulation replications.
B Mathematical Appendix

B.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 $\Theta_I$. This condition ensures that their $QLR$-statistic $\sup_{\theta \in \Theta_I} a_n Q_n(\theta)$ is nondegenerate. Assumption 2.3 (iii) requires the existence of a polynomial function in the distance from $\Theta_I$, 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 B.1 (Consistency and Convergence rate):** Let $t$ be a positive finite constant. Let $t_n = t\kappa_n$ where $\kappa_n$ 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.

B.2 Proof of Auxiliary Lemmas on Duality

In this section, we present definitions and lemmas that are useful for establishing Theorem 3.1.

**Definition B.1 (Level boundedness):** The function $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ is level-bounded if the level sets $\{x : f(x) \leq \alpha\}$ 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 $\omega \in F \in \mathcal{F}$, $P(F) = 1$, then we say $f$ is level bounded almost surely (a.s.).

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

$$s(p, \hat{\Theta}_n(t)) < u \iff \sup_{\theta \in \hat{\Theta}_n(t)} \langle p, \theta \rangle < u$$

$$\iff \langle p, \theta \rangle < u, \quad \forall \theta \in \hat{\Theta}_n(t)$$

$$\iff \hat{\Theta}_n(t) \subseteq \Theta \setminus K_{u,p}$$

$$\iff K_{u,p} \cap \Theta \subseteq \Theta \setminus \hat{\Theta}_n(t)$$

$$\iff \zeta_n(\theta) > t, \quad \forall \theta \in K_{u,p} \cap \Theta$$

$$\iff \inf_{\theta \in K_{u,p} \cap \Theta} \zeta_n(\theta) > t,$$
where the second equivalence follows from the compactness of $\tilde{\Theta}_n(t)$, which is implied by the lower semicontinuity and the level-boundedness of $\zeta_n$, and the last equivalence follows from the properness and the lower semicontinuity of $\zeta_n$, and the compactness of $K_{u,p} \cap \Theta$.

\textbf{Lemma B.1:} 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 $0 < \epsilon < \bar{\epsilon}$,

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

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

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

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 (B.1) holds, we say that $\{\tilde{\zeta}_n\}$ obeys the nonempty limit $\epsilon$-argmin condition. This requires that the sequence $\{\tilde{\zeta}_n\}$ stabilizes in such a way that its $\epsilon$-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).

\textit{Proof of Lemma B.1.} Note first that, under our assumptions, $\tilde{\zeta}_n$ inherits the almost sure properness, lower semicontinuity, and level-boundedness from $\zeta_n$. For any $0 < \epsilon < \bar{\epsilon}$, let $D^*_{n,\epsilon} := \{ (\theta, \lambda) \in R_{u,p} : \tilde{\zeta}_n(\theta, \lambda) < \inf_{R_{u,p}} \zeta_n(\theta, \lambda) + \epsilon \}$ and $D^*_\epsilon := \liminf_{n \to \infty} D^*_{n,\epsilon}$. By hypothesis, $D^*_\epsilon$ is nonempty. For a given $\delta > 0$, let $D^*_\epsilon,\delta$ be an open $\delta$-envelope of $D^*_\epsilon$ defined by $D^*_\epsilon,\delta := \{ (\theta, \lambda) : d((\theta, \lambda), D^*_\epsilon) < \delta \}$.

For any $\delta > 0$, $R_{u,p} \cap D^*_\epsilon,\delta \neq \emptyset$ implies that there exists $N_\epsilon \in \mathbb{N}$ such that $R_{n,u,p} \cap D^*_\epsilon,\delta \neq \emptyset$ for all $n \geq N_\epsilon$ as $R_{n,u,p} \to R_{u,p}$ in the Painlevé-Kuratowski sense (Rockafellar and Wets, 2005, Theorem 4.5). 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 $E_n \neq \emptyset$, $E_{n,\epsilon} \subseteq D^*_\epsilon,\delta$, and $E_{n,\epsilon} \subseteq D_n$, we have $D_n \cap D^*_\epsilon,\delta \neq \emptyset$ 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 $P(D_{n_k} \cap D^*_{n_k,\epsilon} = \emptyset) > 0$. This implies $D_{n_k} \cap D^*_\epsilon,\delta = \emptyset$ for all $k$ with positive probability, which is a contradiction.

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.

[46]
DEFINITION B.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 \( L_\alpha \) 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 \( \{ \tilde{R}_{u_j,p_j} \} \).

**Lemma B.2 (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 \( \epsilon \)-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 \( L_{u_j,p_j}, j = 1, 2, ..., m \) such that

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

**Proof of Lemma B.2.** 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,

\[
P \left( \inf_{R_{u,p}} \tilde{\zeta}_n(\theta,\lambda) - \inf_{R_{u,p}} \tilde{\zeta}_n(\theta,\lambda) \geq \epsilon \right) \leq \epsilon \tag{B.2}
\]

for sufficiently large \( n \). By Lemma 3.1,

\[
Z_n(p,t) < u \iff \inf_{R_{n,u,p}} \zeta_n(\theta + \alpha_n^{1/\gamma}) > t \\
\iff \inf_{R_{n,u,p}} \tilde{\zeta}_n(\theta,\lambda) > t.
\]

Since this holds for any finite \( m \)-tuple \( \{(u_j,p_j)\}_j^m \), we have

\[
P(Z_n(p_1,t) < u_1, ..., Z_n(p_m,t) < u_m) = P \left( \inf_{R_{n,u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, ..., \inf_{R_{n,u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t \right). \tag{B.3}
\]
Note that

\[ P\left( \inf_{R_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t + \epsilon, \ldots, \inf_{R_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t + \epsilon \right) \]

\[ \leq P\left( \max_{1 \leq j \leq m} \inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) - \inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) \geq \epsilon/2 \right) \]

\[ + P\left( \inf_{R_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t + \epsilon/2, \ldots, \inf_{R_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t + \epsilon/2 \right) \]

\[ \leq P\left( \max_{1 \leq j \leq m} \inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) - \inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) \geq \epsilon/2 \right) \]

\[ + P\left( \inf_{R_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \ldots, \inf_{R_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t \right), \]  \quad \text{(B.4)}

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 \( \epsilon \)-contraction \( G^{-\epsilon} := \{ x \in G : \rho(x, G^c) \geq \epsilon \} \), we have \( P(Y_n \in G^{-\epsilon}) \leq P(\rho(X_n, Y_n) \geq \epsilon) + P(X_n \in G) \). Specifically, we used the metric \( \rho(X_n, Y_n) = \max_{1 \leq j \leq m} |X_{j,n} - Y_{j,n}| \) and the open set \( G = (t, \infty)^m \).

Lemma B.1 and (B.2) ensure that the first two terms on the right hand side of (B.4) become arbitrarily small as \( n \) gets large. Therefore,

\[ \lim_{n \to \infty} P\left( \inf_{R_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t + \epsilon, \ldots, \inf_{R_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t + \epsilon \right) \]

\[ \leq \lim_{n \to \infty} P\left( \inf_{R_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \ldots, \inf_{R_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t \right). \]

By letting \( \epsilon \downarrow 0 \), we obtain

\[ \lim_{n \to \infty} P\left( \inf_{R_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \ldots, \inf_{R_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t \right) \]

\[ \leq \lim_{n \to \infty} P\left( \inf_{R_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \ldots, \inf_{R_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t \right) \]

\[ = \lim_{n \to \infty} P(Z_n(p_1, t) < u_1, \ldots, Z_n(p_m, t) < u_m), \]

where the last equality follows from Eq. (B.3). This establishes the claim of the lemma. \( \Box \)

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

For establishing our main result, we require that the local process satisfies additional regularity conditions.

**Assumption B.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 \( \Theta_1 \). (ii) The sequence \( \{ \tilde{\zeta}_n \} \) obeys the nonempty limit \( \epsilon \)-argmin condition, is equi-inf-compact, and weakly epiconverges to a
stochastic process \( \tilde{\zeta} \).

Assumption B.1 (ii) is stronger than strictly necessary. It appears that weak epiconvergence can be replaced by the lower epilimit condition without affecting our conclusions.

We first that the stochastic process \( Z(\cdot, t) \) given in Eq. (3.2) in Theorem 3.1 satisfies \( \{ \omega : Z(p, t) < u \} = \{ \omega : \inf_{\tilde{R}_u,p}(\tilde{\zeta}(\theta, \lambda) > t) \} \) 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 \( g : \theta \mapsto s(p, \tilde{\Lambda}(t, \theta)) \). 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 B.3** (Level-boundedness for parametric optimization): A function \( f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \) with values \( f(x, u) \) is level-bounded in \( x \) locally uniformly in \( u \) if for each \( \tilde{u} \in \mathbb{R}^m \) and \( \alpha \in \mathbb{R} \) there is a neighborhood \( V \in \mathcal{N}(\tilde{u}) \) 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 \( V \in \mathcal{N}(\tilde{u}) \) 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 B.3**: Consider \( \psi(u) := \inf_x f(x, u) \) in the case of a proper, lsc function \( f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \) such that \( f(x, u) \) is level bounded in \( x \) locally uniformly in \( u \). Then, the function \( \psi \) is proper and lsc on \( \mathbb{R}^m \).

**Proof of Lemma B.3**: See Theorem 1.17 in Rockafellar and Wets (2005).

**Lemma B.4**: 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 \( g : \theta \mapsto s(p, \tilde{\Lambda}(t, \theta)) \). Then, there is a representation of \( g \), which is upper semicontinuous (usc) almost surely.

**Proof of Lemma B.4**: First, let \( \delta_A : \mathbb{R}^d \rightarrow \mathbb{R}_+ \) be the optimization theory indicator function that takes \( 0 \) if \( x \in A \) and \( \infty \) otherwise. For each \( \theta \), let \( h(\theta, \lambda) := -\langle p, \lambda \rangle + \delta_{\tilde{\Lambda}(t, \theta)}(\lambda) \) 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 B.3 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 \( \lambda \) locally uniformly in \( \theta \).

By our hypothesis, \( \tilde{\Lambda}(t, \theta) \) is nonempty a.s. for any \( \theta \in \partial \Theta_t \) and \( t \in \mathbb{R}_+ \). Therefore, \( \delta_{\tilde{\Lambda}(t, \theta)} \) is proper, which implies that \( h \) is proper. In the following, using Skorokhod 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 proper lsc functions equipped with a metric that metrizes the topology of epiconvergence is complete and separable (Rockafellar and Wets, 2005). The almost sure epiconvergence of lsc functions \( \{ \tilde{\zeta}_n, n \geq 1 \} \) implies that \( \tilde{\zeta} \) is lsc a.s. (Attouch, 1986, Theorem 2.1). Therefore, the level set \( \tilde{\Lambda}(t, \theta) \) of the lsc function \( \tilde{\zeta}(\theta, \cdot) \) is closed a.s. Note that \( -\langle p, \lambda \rangle \) is continuous and \( \delta_{\tilde{\Lambda}(t, \theta)} \) is lsc by the closedness of \( \tilde{\Lambda}(t, \theta) \). So, \( h \) is lsc a.s.

For each \( p \in \mathbb{S}^{d-1} \) and \( \theta \in H(p, \Theta_t) \), let \( \mathcal{N}(\bar{\theta}) \) be a collection of neighborhoods at \( \bar{\theta} \). Let

\(^{26}\)We follow the convention that \( \sup_{x \in C} f(x) = -\infty \) if \( C \) is an empty set.
\[ \alpha \in \mathbb{R}. \text{Take } \theta \in V \subseteq \mathcal{N}(\bar{\theta}). \text{Define the set} \]

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

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

\[ C = \{ \lambda : \delta_{\Lambda(t, \theta)}(\lambda) = 0, \quad \langle p, \lambda \rangle \leq \alpha \} = \{ \lambda : h(\theta, \lambda) \leq \alpha \}. \]

Therefore, \( h(\theta, \lambda) \) is level bounded in \( \lambda \) locally uniformly in \( \theta \). By Lemma B.3, \( \tilde{g}(\theta) \) is lsc almost surely.

Given the results above, we first 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 \( \Lambda(t, \theta) \subset L_{u, p} \), a.s. For a given \( \theta \in H(p, \Theta_I) \), it is straightforward to show

\[ s(p, \Lambda(t, \theta)) < u \iff \inf_{\lambda \in K_{u, p} \cap L_{u, p}} \tilde{\zeta}(\theta, \lambda) > t, \quad (B.5) \]

using an argument similar to the proof of Lemma 3.1. By the compactness of \( H(p, \Theta_I) \) and Lemma B.4,

\[ \mathcal{Z}(p, t) < u \iff \sup_{\theta \in H(p, \Theta_I)} s(p, \Lambda(t, \theta)) < u \]

\[ \iff s(p, \Lambda(t, \theta)) < u, \quad \forall \theta \in H(p, \Theta_I). \quad (B.6) \]

Combining Eqs. (B.5) and (B.6), we obtain

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

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

\[ \{ \omega : \mathcal{Z}(p_1, t) < u_1, \ldots, \mathcal{Z}(p_m, t) < u_m \} = \left\{ \omega : \inf_{(\theta, \lambda) \in \tilde{R}_{u_1, p_1}} \tilde{\zeta}(\theta, \lambda) > t, \ldots, \inf_{(\theta, \lambda) \in \tilde{R}_{u_m, p_m}} \tilde{\zeta}(\theta, \lambda) > t \right\}. \]

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

**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

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

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


Lemma B.6: Let \( h : [0, \infty) \to [0, \infty) \) be a function such that \( h(0) = 0 \) and \( h \) is continuous at 0. There is \( B_0 \) such that \( B_0 = O_p(1) \). If for all \( x, y \in \mathbb{E} \), \( |\xi_n(x) - \xi_n(y)| \leq B_0 h(\|x - y\|) \), then \( \{ \xi_n \} \) is stochastically equicontinuous.

Proof of Theorem 3.1 (iii). We first show the required conditions for Lemma B.6 using an expansion of the support function based on Lemma B.5. In the following, we extend \( s(\cdot, \Theta_I) \) and \( s(\cdot, \hat{\Theta}_n) \) from \( \mathbb{S}^{d-1} \) to \( \mathbb{R}^d \). Under our assumptions, \( \Theta_I \) 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 (Molchanov, 2005, Theorem F.1). This implies that \( p \mapsto s(p, \Theta_I) \) is strictly continuous, and \( p \mapsto s(p, \hat{\Theta}_n(t)) \) is strictly continuous a.s. Furthermore, the extended support function of a compact convex set is convex (Groemer, 1996, p.20). This ensures that \( p \mapsto s(p, \Theta_I) \) is convex, and \( p \mapsto s(p, \hat{\Theta}_n(t)) \) is convex a.s.

Now, take an open convex set \( O \) such that \( \mathbb{S}^{d-1} \subseteq O \). Let \( p, q \in \mathbb{S}^{d-1} \). Then, by Lemma B.5, for some \( \tilde{p}_n \) and \( \tilde{p} \) on the line segment that connects \( p \) and \( q \), there exist \( \tilde{v}_n \in \partial s(\tilde{p}_n, \hat{\Theta}_n(t)) \) and \( w \in \partial s(\tilde{p}, \Theta_I) \) such that

\[
\begin{align*}
  s(p, \hat{\Theta}_n(t)) - s(q, \hat{\Theta}_n(t)) &= \langle \tilde{v}_n, p - q \rangle \\
  s(p, \Theta_I) - s(q, \Theta_I) &= \langle w, p - q \rangle
\end{align*}
\]

Subtracting (B.8) from (B.7) and multiplying both sides by \( a_n^{1/\gamma} \) yields

\[
Z_n(p, t) - Z_n(q, t) = a_n^{1/\gamma} \langle \tilde{v}_n - w, p - q \rangle.
\]

Note that, Assumption 2.3 (ii) implies \( Z_n(p, t) = O_p(1) \) for any \( p \in \mathbb{S}^{d-1} \). Therefore \( a_n^{1/\gamma} \langle \tilde{v}_n - w, p - q \rangle = Z_n(p, t) - Z_n(q, t) = O_p(1) \) for any \( p, q \in \mathbb{S}^{d-1} \). Since this holds for any \( p \) and \( q \), each component of \( a_n^{1/\gamma} \langle \tilde{v}_n - w \rangle \) must be \( O_p(1) \). Therefore, \( a_n^{1/\gamma} \langle \tilde{v}_n - w \rangle = O_p(1) \).

Applying the Cauchy-Schwarz inequality to (B.9), we obtain

\[
|Z_n(p, t) - Z_n(q, t)| \leq a_n^{1/\gamma} \| \tilde{v}_n - w \| \| p - q \|.
\]

Now, we apply Lemma B.6 with \( B_n = a_n^{1/\gamma} \| \tilde{v}_n - w \| \) and \( h(x) = x \). This ensures that \( \{ Z_n(\cdot, t), n \geq 1 \} \) is stochastically equicontinuous. Thus, \( \{ Z_n(\cdot, t), n \geq 1 \} \) 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 \( Z_n(\cdot, t) \xrightarrow{u.d.} Z(\cdot, t) \).

B.5 Proof of Theorems and Corollaries in Section 3.2

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

\[
U_{n,b}(x, t) := \sum_{k=1}^{N_{n,b}} \left\{ \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}^*(p, t)) \leq x \right\}.
\]

Let \( \epsilon, \delta > 0 \) and \( K \) be the Lipschitz constant of \( \mathcal{Y} \). Suppose \( \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p, t)) \leq x \), \( a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \epsilon/2K \), \( d_H(\hat{\Psi}_n, \Psi_0) \leq \delta \), and \( \sup_{\| p-p' \| \leq \delta} |Z_{n,b,k}(p, t) - Z_{n,b,k}(p)| \leq \epsilon/2K \).

[52]
Therefore, \(
\sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) - \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}^*(p,t)) \mid
\)
\[
\leq \mid \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) - \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) \mid + \sup_{p \in \Psi_0} \sup_{p \in \mathbb{R}^{d-1}} \mid \mathcal{Y}(Z_{n,b,k}(p,t)) - \mathcal{Y}(Z_{n,b,k}^*(p,t)) \mid
\]
\[
\leq \mid \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) - \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) \mid + \sup_{p \in \mathbb{R}^{d-1}} \mid \mathcal{Y}(Z_{n,b,k}(p,t)) - \mathcal{Y}(Z_{n,b,k}^*(p,t)) \mid
\]

(B.11)

Let \( \hat{p}_n \in \arg \max_{p \in \Psi_n} \mathcal{Y}(Z_{n,b,k}(p,t)) \), which is well defined by the compactness of \( \hat{\Psi}_n \) and the continuity of the map \( p \mapsto \mathcal{Y}(Z_{n,b,k}(p,t)) \). Let \( \Pi_{\Psi_0} \hat{p}_n \) be the projection of \( \hat{p}_n \) on \( \Psi_0 \) and note that \( \| \hat{p}_n - \Pi_{\Psi_0} \hat{p}_n \| \leq d_H(\hat{\Psi}_n, \Psi_0) \leq \delta \). We thus obtain,
\[
\sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) - \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) \leq \mathcal{Y}(Z_{n,b,k}(\hat{p}_n, t)) - \mathcal{Y}(Z_{n,b,k}(\Pi_{\Psi_0} \hat{p}_n, t)) \leq \sup_{\| p - p' \| \leq \delta} \mid \mathcal{Y}(Z_{n,b,k}(p,t)) - \mathcal{Y}(Z_{n,b,k}(p', t)) \mid
\]

A similar argument ensures
\[
\sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) - \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) \leq \sup_{\| p - p' \| \leq \delta} \mid \mathcal{Y}(Z_{n,b,k}(p,t)) - \mathcal{Y}(Z_{n,b,k}(p', t)) \mid
\]

Therefore,
\[
\mid \sup_{p \in \Psi_n} \mathcal{Y}(Z_{n,b,k}(p,t)) - \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}(p,t)) \mid \leq \sup_{\| p - p' \| \leq \delta} \mid \mathcal{Y}(Z_{n,b,k}(p,t)) - \mathcal{Y}(Z_{n,b,k}(p', t)) \mid
\]
\[
\leq K \sup_{\| p - p' \| \leq \delta} \mid Z_{n,b,k}(p,t) - Z_{n,b,k}(p', t) \mid \leq \frac{\epsilon}{2}. \quad \text{(B.12)}
\]

by the Lipschitz continuity of \( \mathcal{Y} \) and the hypothesis. Furthermore, the Lipschitz continuity of \( \mathcal{Y} \) and the hypothesis that \( a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \epsilon/2K \) ensure
\[
\sup_{p \in \mathbb{R}^{d-1}} \mid \mathcal{Y}(Z_{n,b,k}(p,t)) - \mathcal{Y}(Z_{n,b,k}^*(p,t)) \mid \leq K \sup_{p \in \mathbb{R}^{d-1}} a_b^{1/b} \mid s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I) \mid \leq \frac{\epsilon}{2}. \quad \text{(B.13)}
\]

Combining Eqs. (B.11)-(B.13) yields \( \sup_{p \in \Psi_n} \mathcal{Y}(Z_{n,b,k}(p,t)) - \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}^*(p,t)) \mid \leq \epsilon \).

Therefore, \( \sup_{p \in \Psi_0} \mathcal{Y}(Z_{n,b,k}^*(p,t)) \leq x + \epsilon \). Let
\[
E_{n,b}(t, \epsilon, \delta) := \{ \omega \in \Omega : a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \epsilon/2K, \quad d_H(\hat{\Psi}_n, \Psi_0) \leq \delta \},
\]
\[
\sup_{\| p - p' \| \leq \delta} \mid Z_{n,b,k}(p,t) - Z_{n,b,k}(p,t) \mid \leq \epsilon/2K.
\]

The arguments above show:
\[
\bar{F}_{n,b}(x,t) 1_{E_{n,b}(t, \epsilon, \delta)} \leq U_{n,b}(x + \epsilon, t). \quad \text{(B.14)}
\]
Now, suppose \( \sup_{p \in \Psi_0} \Psi(Z_{n,b,k}^*(p,t)) \leq x - \epsilon \) and that \( d_{H}^{1/\gamma} d_{H}^{1/\gamma}(\hat{\Theta}_n(t), \Theta_I) \leq \epsilon/2K, \), \( d_{H}(\hat{\Psi}_n, \Psi_0) \leq \delta \), and \( \sup_{|[p-p']| \leq \delta} |Z_{n,b,k}(p,t) - Z_{n,b,k}(p)| \leq \epsilon/2K \). Following the same argument as before, we obtain \( \sup_{p \in \Psi_n} \Psi(Z_{n,b,k}^*(p,t)) \leq \epsilon \). Thus, it follows that \( \sup_{p \in \Psi_n} \Psi(Z_{n,b,k}(p,t)) \leq \epsilon \). Therefore,

\[
U_{n,b}(x - \epsilon, t) 1_{E_{n,b}(t, \epsilon, \delta)} \leq \hat{F}_{n,b}(x, t) 1_{E_{n,b}(t, \epsilon, \delta)}, \tag{B.15}
\]

Since (B.14) and (B.15) hold for any \( \epsilon, \delta > 0 \) and \( P(E_{n,b}(t, \epsilon, \delta)) \to 1 \) as \( n \to \infty \) and \( b \to \infty \) by Theorem 2.1 (i), the assumption that \( d_{H}(\hat{\Psi}_n, \Psi_0) = o_p(1) \), and the stochastic equicontinuity of \( \{Z_{n,b,k}(\cdot, t)\} \) as proved in the proof of Theorem 3.1 (iii), we have

\[
U_{n,b}(x - \epsilon, t) \leq \hat{F}_{n,b}(x, t) \leq U_{n,b}(x + \epsilon, t), \tag{B.16}
\]

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

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

\[
F(x - \epsilon, t) - \epsilon \leq \hat{F}_{n,b}(x, t) \leq F(x + \epsilon, t) + \epsilon,
\]

with probability tending to 1 for any \( \epsilon > 0 \). Now, let \( \epsilon \downarrow 0 \) so that \( x \pm \epsilon \) are continuity points of \( F^{-}(\cdot, P) \). 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 3.1. (i) Let \( \hat{F}_{n,-}^{*}(\cdot, t) \) be the empirical cdf of \( T_{n,-}^{*}(t) \). Similarly, let \( F^{*}(\cdot, t) \) be the cdf of \( \sup_{p \in \Psi_{-1}} \{ -Z(p, t) \} \). Let \( \Psi(x) = \{-x\} \) and \( \Psi_0 = \Psi_n = \mathbb{S}^{d-1} \). By Theorem 3.2, it follows that \( \hat{F}_{n,-}^{*}(x, t) - F^{*}(x, t) = o_p(1) \) at every continuity point of \( F^{*}(\cdot, t) \). The consistency of \( \hat{c}_{n,b,1-\alpha}(t) \) then follows from Lemma 11.2.1 in Lehmann and Romano (2005).

Under the null hypothesis, we have \( T_{n,-}^{*}(t) \leq a_n^{1/\gamma} \tilde{d}_{H}(\Theta_I, \hat{\Theta}_n(t)), \) and \( \tilde{d}_{H}(\Theta_I, \hat{\Theta}_n(t)) \) converges in distribution to \( F^{*}(\cdot, t) \) by Theorem 3.1, Lemma A.1 in BM, and the continuous mapping theorem. By the results above and by Corollary 11.2.3 in Lehmann and Romano (2005), we have

\[
\limsup_{n \to \infty} P(T_{n,-}^{*}(t) > \hat{c}_{n,b,1-\alpha}(t)) \leq \lim_{n \to \infty} P(a_n^{1/\gamma} \tilde{d}_{H}(\Theta_I, \hat{\Theta}_n(t)) > \hat{c}_{n,b,1-\alpha}(t)) = 1 - F(c_{1-\alpha}(t), t) = \alpha.
\]

The proof of part (ii) is very similar to the proof of Corollary 2.5 in BM.

Proof of Theorem 3.3. The proof is similar to the proof of Proposition 2.7 in BM.

Proof of Lemma 3.2. First, \( c_{1-\alpha}(t_{1-\alpha}) = 0 \) follows from the definition of \( t_{1-\alpha} \). For the conclusion of the lemma, it suffices to show that \( P(\sup_{p \in \mathbb{S}^{d-1}} \{ -Z(p, t) \} \leq x) \) is non-decreasing in \( t \) for each \( x \). As this is a distributional property of the process \( Z(p, t) \), it suffices to show...
that the statement above holds for the following representation:

\[-Z(p, t) = - \sup_{\theta \in H(p, \Theta_t)} \sup_{\lambda \in \{\lambda : \tilde{\zeta}(\theta, \lambda) \leq t\}} \langle p, \lambda \rangle.\]

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 \(p \in S^{d-1}\), 

\(-Z(p, t)\) is non-increasing in \(t\). This implies that \(\sup_{p \in S^{d-1}} \{-Z(p, t)\}_+\) is non-increasing in \(t\) for any \(\omega\). Thus, \(P(\sup_{p \in S^{d-1}} (-Z(p, t))_+ \leq x)\) is non-decreasing in \(t \in [0, t_{1-\alpha}^*]\) for each \(x\).

\(\square\)

We use the following lemma to prove Theorem 3.4.

**Lemma B.7:** Suppose the conditions of Theorem 3.4 hold. Then, for any \(\alpha \in (0, 1)\) and 0 \(\leq t < t' \leq t_{1-\alpha}^*\), \(c_{1-\alpha}'(t) - c_{1-\alpha}'(t') = \mu(t') - \mu(t)\).

**Proof of Lemma B.7.** First, \(c_{1-\alpha}'(t)\) can be written as

\[c_{1-\alpha}'(t) = \inf \left\{ x : P \left( \sup_{p \in S^{d-1}} \{-Z(p, t)\}_+ \leq x \right) \geq 1 - \alpha \right\}\]

\[= \inf \left\{ x : P \left( \sup_{p \in S^{d-1}} \{\mu(t') - \mu(t) - \mu(t') - Z^*(p)\}_+ \leq x \right) \geq 1 - \alpha \right\}. \quad (B.17)\]

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

\[P \left( \sup_{p \in S^{d-1}} \{\mu(t') - \mu(t) - \mu(t') - Z^*(p)\}_+ \leq x \right) \]

\[= P \left( \sup_{p \in S^{d-1}} \{\Delta(t, t') - Z(p, t')\}_+ \leq x \right) \]

\[= P \left( \sup_{p \in S^{d-1}} \{-Z(p, t')\}_+ \leq x - \Delta(t, t') \right). \quad (B.18)\]

Substituting Eq. (B.18) into Eq. (B.17) yields

\[c_{1-\alpha}'(t) = \inf \left\{ x : P \left( \sup_{p \in S^{d-1}} \{-Z(p, t')\}_+ \leq x - \Delta(t, t') \right) \geq 1 - \alpha \right\} \]

\[= c_{1-\alpha}'(t') + \Delta(t, t'). \quad \square\]
Proof of Theorem 3.4. By Theorem in 1.1.12 in Li, Ogura, and Kreinovich (2002),
\[
a_n^{1/\gamma} d_H \left( \hat{\Theta}_n^{*,b,1-\alpha}(t), \hat{\Theta}_n(t_{1-\alpha}) \right) = a_n^{1/\gamma} \sup_{p \in S^d-1} |s(p, \hat{\Theta}_n(t)) + \hat{c}_n^{*,b,1-\alpha}(t) - s(p, \hat{\Theta}_n(t_{1-\alpha}))|
\]
\[
= \sup_{p \in S^d-1} |a_n^{1/\gamma}[s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I)] - a_n^{1/\gamma}[s(p, \hat{\Theta}_n(t_{1-\alpha})) - s(p, \Theta_I)] + \hat{c}_n^{*,b,1-\alpha}(t)|
\]
\[
= \sup_{p \in S^d-1} |Z_n(p, t) - Z_n(p, t_{1-\alpha})| + c_{1-\alpha}^*(t) + o_p(1)|
\]
\[
\leq (1) \sup_{p \in S^d-1} |\mu(t) - \mu(t_{1-\alpha})| - (c_{1-\alpha}^*(t_{1-\alpha}) - c_{1-\alpha}^*(t)) + o_p(1)|
\]
\[
= o_p(1),
\]
where we used the fact that \(c_{1-\alpha}^*(t_{1-\alpha}) = 0\) in equality (1), and the last equality follows from Lemma B.7.

For (ii), the result immediately follows from Theorem 3.4 and the triangle inequality:
\[
d_H \left( \hat{\Theta}_n^{*,b,1-\alpha}(t), \hat{\Theta}_n^{*,b,1-\alpha}(t') \right) \leq d_H \left( \hat{\Theta}_n^{*,b,1-\alpha}(t), \hat{\Theta}_n(t_{1-\alpha}) \right) + d_H \left( \hat{\Theta}_n(t_{1-\alpha}), \hat{\Theta}_n(t_{1-\alpha}) \right).
\]

The conclusion of Theorem 3.5 can be established by applying a functional \(\delta\)-method. For this, we need a suitable differentiability concept of the map \(s(\cdot, \Theta_I) \mapsto \sup_{p \in S^d-1} \langle p, \theta_0 \rangle - s(p, \Theta_I)\). The following definition is based on Shapiro (1991)

**Definition B.5:** Let \(X\) and \(Y\) be normed vector spaces. A map \(g : X \to Y\) is said to be Hadamard directionally differentiable at \(\mu\) if for every sequence \(\{t_n\}\) of positive numbers converging to 0 and any sequence \(\{x_n\}\) converging to \(x\), the limit
\[
\hat{g}(x) = \lim_{n \to \infty} \frac{g(\mu + t_n x_n) - g(\mu)}{t_n}
\]
exists. If \(\hat{g}\) is linear in \(x\), then \(g\) is said to be Hadamard differentiable at \(\mu\).

The following lemma is useful for establishing the Hadamard directional differentiability of the map \(s(\cdot, \Theta_I) \mapsto \sup_{p \in S^d-1} \langle p, \theta_0 \rangle - s(p, \Theta_I)\).

**Lemma B.8:** Let \(S\) be a compact subset of a finite dimensional Euclidean space. Let \(B \equiv C(S)\) be the space of continuous functions on \(S\). For a given \(g \in B\), let \(\hat{\phi}_g : B \to \mathbb{R}\) be defined pointwise by \(\hat{\phi}_g(x) := \sup_{p \in S} g(p) - x(p)\). Then, for any \(x \in B\), \(\hat{\phi}_g\) is Hadamard directionally differentiable at \(x\), and its directional derivative \(\hat{\phi}_g : B \to \mathbb{R}\) is given pointwise by
\[
\hat{\phi}_g(y) := \sup_{p \in \Psi(g-y)} -y(p),
\]
where for each \(z \in B\), \(\Psi(z) := \text{argmax}_{p \in S} z(p)\). Furthermore, if \(\Psi(g-x)\) is singleton-valued, \(\hat{\phi}_g\) is Hadamard differentiable at \(x\).
Proof of Lemma B.8. The proof is a modification of Theorem 3.1 in Shapiro (1991). First, we show that $\dot{\phi}_g$ is a continuous functional. Let $\{y_n\}$ be a sequence such that $y_n \to y$ for some $y \in \mathcal{B}$. Note that $\Psi(g-x)$ is nonempty and compact by Theorem 17.31 in Aliprantis and Border (2006). Since $-y_n$ converges uniformly to $-y$ on $\Psi(g-x)$, $\max_{p \in \Psi(g-x)} \Psi(g-x) = -y_n(p) \to \max_{p \in \Psi(g-x)} \Psi(g-x) = -y(p)$. Since the choice of $y$ was arbitrary, $\dot{\phi}_g$ is continuous at every point.

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

$$w(p) \geq f'_p(-w)$$

for all $w \in \mathcal{B}$. Furthermore, since $f'_p \neq -e_p$, there exists a $w \in \mathcal{B}$ such that $w(p) > f'_p(-w)$. Let $w' := y - w$. Then, $w'(p) = y(p) - w(p) < y(p) - f'_p(-w) = y(p) + f'_p(w) = y(p) + f'_p(y - w')$, which contradicts (B.21). Therefore, $-e_p$ is the unique element of $\partial f_p(y)$.

Fix $y \in \mathcal{B}$. We note that $S$ is a compact subset of a Hausdorff space and that $f_p$ is continuous for every $p \in S$. Furthermore, for any $p \in S$ and $\{p_n\} \subset S$ such that $p_n \to p$, it follows that $f_{p_n}(y) = y(p_n) \to y(p) = f_p(y)$ by the continuity of $y$. Therefore $p \mapsto f_p(y)$ is continuous at every $y$. Now the conditions of Theorem 2.4.18 in Zalinescu (2002) are satisfied. This implies that the subdifferential of $\dot{\phi}_g$ at $y$ takes the form:

$$\partial \dot{\phi}_g(y) = \mathcal{C}\mathcal{O}(\cup_{f_p = \sup_{p \in S} f_p \partial f_p(y)}) = \mathcal{C}\mathcal{O}(\cup_{p \in \Psi(g-x)} \{-e_p\})$$

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

$$\tilde{e}_{n_k}(y) = \lambda_{n_k}(-y(p_{n_k})) + (1 - \lambda_{n_k})(-y(p'_{n_k})) \to \lambda^*(-y(p^*)) + (1 - \lambda^*)(-y(p'^*)) = \lambda^*(-e_{p^*}(y)) + (1 - \lambda^*)(-e_{p'^*}(y)) \in C,$$

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

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

\[
\dot{\phi}_g^G(y) = \sup_{\phi'_g \in \partial \phi_g} \phi'_g(y) \\
= \sup_{\lambda \in [0, 1]} \sup_{p, p' \in \Psi(g-x)} \lambda(-y(p)) + (1 - \lambda)(-y(p')). \tag{B.22}
\]

Now suppose that \( \text{argmax}_{\Psi(g-x)} y(p) = \{ \bar{p} \} \) for some \( \bar{p} \in \Psi(g-x) \), then the right hand side of (B.22) is equal to \(-y(\bar{p})\). Therefore, in this case \( \dot{\phi}_g^G(y) = -y(\bar{p}) = \sup_{p \in \Psi(g-x)} -y(p) \).

Similarly if \( \text{argmax}_{\Psi(g-x)} y(p) \) is not a singleton, then again the right hand side of (B.22) is equal to \(-y(\bar{p})\) for some \( \bar{p} \in \Psi(g-x) \) because \(-y(\bar{p}) \geq \lambda(-y(\bar{p})) + (1 - \lambda)(-y(p)) \) for all \( \lambda \in [0, 1] \) and \( p \), and equality holds only if \( p \) is also in \( \text{argmax}_{\Psi(g-x)} -y(p) \). Therefore, it follows again that \( \dot{\phi}_g^G(y) = \sup_{p \in \Psi(g-x)} -y(p) \). This establishes that \( \dot{\phi}_g \) in (B.20) is the Gateaux directional derivative of \( \phi_g \).

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

\[
|h_n(y) - h_n(y')| = |\phi_g(x + t_n y) - \phi_g(x + t_n y')|/t_n - \dot{\phi}_g^G(y - y')| \\
\leq \|x + t_n y - (x + t_n y')\|/t_n + \|y - y'\| = 2\|y - y'\|.
\]

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

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

**Lemma B.9:** Let \( m \in \mathbb{N} \). Let \( D \subseteq \mathbb{R}^m \) be a compact convex set with a nonempty interior. Let \( K_0 \) be a nonempty closed convex subset of \( D \) and \( \{ \tilde{K}_n \} \) be a sequence of measurable closed convex subsets of \( D \). Given a positive sequence \( \{ \tau_n \} \) such that \( \tau_n \to \infty \), let \( \mathcal{W}_n := \tau_n(s(\cdot, \tilde{K}_n) - s(\cdot, K_0)) \). Given \( x_0 \in \partial K_0 \), let

\[
S_{n, \theta_0} := \tau_n \sup_{p \in \mathbb{S}^{m-1}} \{ \langle p, x_0 \rangle - s(p, \tilde{K}_n) \} + \tag{B.23}
\]

and

\[
L_0 := \arg \max_{p \in \mathbb{S}^{m-1}} \langle p, x_0 \rangle - s(p, K_0) \tag{B.24}
\]

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

\[
S_{n, \theta_0} \overset{d}{\to} \sup_{p \in L_0} \{-\mathcal{W}(p)\} +. \tag{B.25}
\]

[58]
Proof of Lemma B.9. We first note that \( x_0 \in \partial K_0 \) implies \( \sup_{p \in S^{m-1}} \langle p, x_0 \rangle - s(p, K_0) = 0 \). Let \( \phi_{x_0} : C(S^{m-1}) \rightarrow \mathbb{R} \) be defined pointwise by \( \phi_{x_0}(f) := \sup_{p \in S^{m-1}} \langle p, x_0 \rangle - f(p) \). Now the statistic can be written as

\[
S_{n,x_0}^+ = \max\{ \tau_n(\phi_{x_0}(s(\cdot, \hat{K}_n))) - \phi_{x_0}(s(\cdot, K_0)), 0 \}. \tag{B.26}
\]

By Lemma B.8, \( \phi_{x_0} \) is Hadamard directionally differentiable at \( s(\cdot, K_0) \) with Hadamard directional derivative \( \hat{\phi}_{x_0}(y) = \sup_{p \in L_0} -y(p) \). This and the assumption that \( W_n \overset{u.d.}{\rightarrow} \mathcal{W} \) ensure the conditions of Theorem 2.1 in Shapiro (1991). It follows that

\[
\tau_n(\phi_{x_0}(s(\cdot, \hat{K}_n))) - \phi_{x_0}(s(\cdot, K_0)) \overset{d}{\rightarrow} \sup_{p \in L_0} -W(p). \tag{B.27}
\]

The conclusion of the Lemma now follows from (B.26), (B.27), and the continuous mapping theorem. \( \square \)

Proof of Theorem 3.5. Let \( t \geq 0 \). We apply Lemma B.9 with \( D = \Theta, K_0 = \Theta_1, \hat{K}_n = \hat{\Theta}_n(t) \), and \( \tau_n = a_n^{\gamma} \). Under our hypothesis, Theorem 3.1 holds. The conclusion of Theorem 3.1 ensures that \( W_n = a_n^{1/\gamma}(s(p, \hat{\Theta}_n(t)) - s(p, \Theta_0)) \) converges weakly to a tight limit \( \mathcal{W} = \mathcal{Z}(\cdot, t) \).

By setting \( x_0 = \theta_0 \) and \( L_0 = \Psi_0 \), Lemma B.9 ensures

\[
T_{n,\theta_0} \overset{d}{\rightarrow} \sup_{p \in \Psi_0} \{-Z(p, t)\}_.
\]

This completes the proof. \( \square \)

In order to ensure the validity of subsampling, the following two lemmas are useful.

**Lemma B.10:** Let \( m \in \mathbb{N} \). Let \( D \subset \mathbb{R}^m \) be a compact convex set with a nonempty interior. Let \( K_0 \) be a nonempty closed convex subset of \( D \) and \( \{\hat{K}_n\} \) be a sequence of measurable closed convex subsets of \( D \) such that

\[
d_H(\hat{K}_n, K_0) = O_p(a_n^{-1/\gamma}), \tag{B.28}
\]

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

\[
L_0 := \arg \max_{p \in S^{m-1}} \langle p, x_0 \rangle - s(p, K_0). \tag{B.29}
\]

Given a positive sequence \( \{\kappa_n\} \), let

\[
\hat{L}_n := \{p \in S^{m-1} : \langle p, x_0 \rangle - s(p, \hat{K}_n) \geq \sup_{p' \in S^{m-1}} [(\langle p', x_0 \rangle - s(p', \hat{K}_n)] - \kappa_n/a_n^{1/\gamma}\}. \tag{B.30}
\]

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

**Proof of Lemma B.10.** In order to prove the claim, we invoke Theorem 3.1 in Chernozhukov, Hong, and Tamer (2007). First note that \( S^{m-1} \) is nonempty and compact. Let \( Q \) and \( Q_n \) be defined pointwise by \( Q(p) := [s(p, K_0) - \langle p, x_0 \rangle] - \inf_{p' \in S^{m-1}} [s(p', K_0) - \langle p', x_0 \rangle] \) and \( Q_n(p) := [s(p, \hat{K}_n) - \langle p, x_0 \rangle] - \inf_{p' \in S^{m-1}} [s(p', \hat{K}_n) - \langle p', x_0 \rangle] \). Note that \( \inf_{p' \in S^{m-1}} [s(p', K_0) - \langle p', x_0 \rangle] \) is finite, and \( \inf_{p' \in S^{m-1}} [s(p', \hat{K}_n) - \langle p', x_0 \rangle] \) is finite almost surely because the objective functions
are continuous (almost surely for the latter) and \( S^{m-1} \) is compact. Note also that \( L_0 \) and \( \hat{L}_n \) can be equivalently written as

\[
L_0 = \arg\min_{p \in S^{m-1}} Q(p) = \{ p \in S^{m-1} : Q(p) = 0 \},
\]

\[
\hat{L}_n = \{ p \in S^{m-1} : Q_n(p) \leq \kappa_n/a_n^{1/\gamma} \}.
\]

Since \( s(p, K_0) \) is continuous by Theorem 1.1.12 in Li, Ogura, and Kreinovich (2002), \( Q \) is continuous. Similarly, since \( s(p, K_n) \) is continuous in \( p \) for each \( \omega \in \Omega \) and measurable for each \( p \), \( s(p, K_n) \) is jointly measurable by Lemma 4.51 in Aliprantis and Border (2006). Furthermore, \( \inf_{p' \in S^{m-1}} [s(p', K_n) - \langle p', x_0 \rangle] \) is measurable by Theorem 2.27 (i) in Molchanov (2005). Thus, \( Q_n \) is jointly measurable.

By (B.28) and Theorem 1.1.12 in Li, Ogura, and Kreinovich (2002), \( s(\cdot, \hat{K}_n) - s(\cdot, K_0) = o_p(1) \) uniformly. Therefore, for any \( \epsilon > 0 \), \( \sup_{p \in S^{m-1}} |s(\cdot, \hat{K}_n) - s(\cdot, K_0)| < \epsilon/2 \) with probability approaching 1. This implies that, with probability approaching 1,

\[
\sup_{p \in S^{m-1}} |Q_n(p) - Q(p)| \leq \sup_{p \in S^{m-1}} |s(\cdot, \hat{K}_n) - s(\cdot, K_0)| + \inf_{p' \in S^{m-1}} [s(p', K_n) - \langle p', x_0 \rangle] - \inf_{p' \in S^{m-1}} [s(p', K_0) - \langle p', x_0 \rangle] < \frac{\epsilon}{2} + |s(\cdot, K_0) + \epsilon/2 - \langle p', x_0 \rangle| - \inf_{p' \in S^{m-1}} [s(p', K_0) - \langle p', x_0 \rangle] = \epsilon.
\]

Thus \( Q_n - Q = o_p(1) \) uniformly. Furthermore, uniformly over \( L_0 \), \( Q_n(p) = [s(p, K_0) + O_p(a_n^{-1/\gamma}) - \langle p, x_0 \rangle] - \inf_{p' \in S^{m-1}} [s(p', K_0) + O_p(a_n^{-1/\gamma}) - \langle p', x_0 \rangle] = O_p(a_n^{-1/\gamma}) \), where the first equality follows from (B.28), and the second equality follows from the construction of \( L_0 \). Hence, under our hypothesis, \( \kappa_n \geq \sup_{p \in L_0} a_n^{1/\gamma} Q_n(p) \) with probability approaching 1. Therefore, all required conditions for Theorem 3.1 (1) in Chernozhukov, Hong, and Tamer (2007) are satisfied. This ensures the claim of the lemma.

**Proof of Corollary 3.2.** (i) Let \( \hat{F}_n \rightarrow (x, \theta_0, t) \) be the empirical cdf of \( T_{n, \theta_0} \rightarrow (t) \). Similarly, let \( F^\rightarrow (x, \theta_0, t) \) be the cdf of \( \sup_{p \in \Psi_0} \{-Z(p, t)\}_+ \). Note that \( d_H(\Psi_0, \hat{\Psi}_n) = o_p(1) \) by Theorem B.1 and Lemma B.10 applied with \( K_0 = \Theta_I, \hat{K}_n = \hat{\Theta}_n(t), L_0 = \Psi_0, \) and \( \hat{L}_n = \hat{\Psi}_n \). Thus, by Theorem 3.2 with \( Y(x) = \{-x\}_+, \Psi_0 = \arg\max_{p, \theta_0} \langle p, \theta_0 \rangle - s(p, \Theta_I), \) and \( \hat{\Psi}_n \) as in (3.8), \( \hat{F}_n \rightarrow (x, \theta_0, t) - F^\rightarrow (x, \theta_0, t) = o_p(1) \) at each continuity point of \( F^\rightarrow (\cdot, \theta_0, t) \). Thus, \( \hat{c}_{n,b,1-\alpha}^\rightarrow (\theta_0, t) = \hat{c}_{n,b,1-\alpha}^\rightarrow (\theta_0, t) + o_p(1) \) by and Lemma 11.2.1 in Lehmann and Romano (2005). Note that if \( \theta_0 \in \partial \Theta_I, T_{n, \theta_0} \rightarrow (t) \) converges in distribution to \( F^\rightarrow (x, \theta_0, t) \) by Theorem 3.5. By Corollary 11.2.3 in Lehmann and Romano (2005), \( \lim_{n \to \infty} P(T_{n, \theta_0} \rightarrow (p_0, t) \leq \hat{c}_{n,b,1-\alpha}^\rightarrow (\theta_0, t)) = F^\rightarrow (c_{1-\alpha}^\rightarrow (p_0, t), t) = 1 - \alpha \). If \( \theta \in \Theta_I \),

\[
T_{n, \theta_0} \rightarrow (t) = \sup_{p \in \Psi_0^I} \{-Z_n(p, t) + a_n^{1/\gamma} (\langle p, \theta_0 \rangle - s(p, \Theta_I))\}_+ \overset{P}{\to} 0.
\]

(B.31)
Therefore,
\[
\limsup_{n \to \infty} P(T_{n, \theta_0}(t) > \hat{c}_{n, 1-\alpha}(\theta_0, t)) \leq \lim_{n \to \infty} P\left(\sup_{p \in \mathbb{S}^l} \{-Z_n(p, t)\} > \hat{c}_{n, 1-\alpha}(\theta_0, t)\right)
\]
\[
= F^{-}(c_{1-\alpha}(\theta_0, t), t) = 1 - \alpha.
\]

Part (ii) follows from the fact that
\[
T_{n, \theta_0}(t) = \sup_{p \in \mathbb{S}^l} \{-Z_n(p, t) + a_n^{1/\gamma}(\langle p, \theta_0 \rangle - s(p, \Theta_I))\} \overset{P}{\to} \infty,
\]
under a fixed alternative and that \(\hat{c}_{n, 1-\alpha}(\theta_0, t) = O_p(1).\)

\[\square\]

**Proof of Theorem 3.6.** The result simply follows from Corollary 3.2 (i) and the equivalence
\[
\theta_0 \in \hat{\Theta}_{n, 1-\alpha}(t) \iff T_{n, \theta_0} \leq \hat{c}_{n, 1-\alpha}(\theta_0, t).
\]

\[\square\]

### B.6 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 B.6 (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 \(x_1, \ldots, x_m \in B\) and open neighborhoods \(V(x_1), \ldots, V(x_m)\) of \(x_1, \ldots, x_m\) such that
\[
B \subset \bigcup_{i=1}^{m} V(x_i)
\]
and
\[
\limsup_{n \to \infty} P\left(\bigcup_{i=1}^{m} \left\{\inf_{y \in V(x_i)} \xi_n(y) \leq \min\{\epsilon, \xi_n(x_i) - \epsilon\}\right\}\right) < \delta.
\]

**Lemma B.11 (Knight, 1999, Theorem 2):** Let \(\{\xi_n, n \geq 1\}\) be a stochastically e-lsc sequence of functions and \(\xi\) 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

$$\begin{align*}
P \left( \sup_{\Theta} \left| \varphi(\hat{E}_n(m_{j,\theta}, \hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}), W(\theta)) \right| > \epsilon \right) \\
\leq P \left( \sup_{\Theta} \left| \varphi(\hat{E}_n(m_{j,\theta}, \hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}), \hat{W}_n(\theta)) \right| + \sup_{\Theta} \left| \varphi(E(m_{j,\theta}), \hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}), W(\theta)) \right| > \epsilon \right) \\
\leq P \left( \sup_{\Theta} \left| \varphi(\hat{E}_n(m_{j,\theta}, \hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}), \hat{W}_n(\theta)) \right| > \epsilon/2 \right) \\
+ P \left( \sup_{\Theta} \left| \varphi(E(m_{j,\theta}), \hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}), W(\theta)) \right| > \epsilon/2 \right) \\
\leq P \left( \sup_{\Theta} L_1 h_1 \left( \left\| \hat{E}_n(m_{j,\theta}) - E(m_{j,\theta}) \right\| \right) > \epsilon/2 \right) \\
+ P \left( \sup_{\Theta} L_2 h_2 \left( \max_{i,j} \left| \hat{W}_{n,ij}(\theta) - W_{ij}(\theta) \right| \right) \right) > \epsilon/2 \right) \\
\leq P \left( \sup_{\Theta} \left\| \hat{E}_n(m_{j,\theta}) - E(m_{j,\theta}) \right\| > L_1^{-1} h_1^{-1} (\epsilon/2) \right) \\
+ P \left( \sup_{\Theta} \max_{i,j} \left| \hat{W}_{n,ij}(\theta) - W_{ij}(\theta) \right| > L_2^{-1} h_2^{-1} (\epsilon/2) \right) \\
\leq \epsilon
\end{align*}$$

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 \( \sup_{\Theta} a_n Q_n(\theta) = O_p(1) \). Therefore, Assumptions 2.3 (ii) holds. Now, let \( \eta > 0 \) be such that \( \sup_{\Theta} \max_{i,j} \left| \hat{W}_{n,ij}(\theta) \right| \leq \eta < \infty, wp \rightarrow 1 \). We can write

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

uniformly over \( \Theta_f \) \( wp \rightarrow 1 \) by Assumption 4.3 (ii-d). We thus have \( Q_n(\theta) = 0 \) on \( \Theta_f^{-\epsilon_n} \) 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 \( \inf_{\Theta_f} \min_{i,j} |\hat{W}_{n,ij}(\theta)| \geq \eta' \), \( wp \rightarrow 1 \). First, we write

$$\begin{align*}
n^{\gamma/2} Q_n(\theta) &= \varphi(G_n(m_\theta) + \sqrt{n}E(m_\theta), \hat{W}_n(\theta)) \\
&\geq C_2 \left\| G_n(m_\theta) + \sqrt{n}E(m_\theta) \right\| \gamma \\
&\geq C_2 \left\| \sqrt{n}E(m_\theta) \right\| \gamma \frac{\left\| G_n(m_\theta) + \sqrt{n}E(m_\theta) \right\|}{\left\| \sqrt{n}E(m_\theta) \right\|}.
\end{align*}$$

Now, by Assumption 4.3 (ii-b), we have \( \left\| \sqrt{n}E(m_\theta) \right\| \gamma \geq C_1 n^{\gamma/2} \min \{ d(\theta, \Theta_f), \delta \} \gamma \) on \( \Theta \) for
some $C_1 > 0$ and $δ > 0$. Therefore, for any $ε > 0$, we can choose $(κ_ε, n_ε)$ so that for any $n ≥ n_ε$ with probability at least $1 - ε$,

$$n^{γ/2}Q_n(θ) ≥ C_2C_1n^{γ/2} \min\{d(θ, Θ_I), δ\}^γ,$$

uniformly over $\{θ ∈ Θ : d(θ, Θ_I) ≥ (κ_ε/n^{γ/2})^{1/γ}\}$, which follows by $∥y + x∥_+ / ∥x∥_+ → 1$ as $∥x∥_+ → ∞$ for any $y ∈ ℝ^J$ and by $\sup_{θ ∈ Θ}∥G_n(θ)∥ = O_p(1)$ by the $P$-Donsker property.

Note that the lower semicontinuity $ζ_ε$ follows the continuity in $θ$ of $ϕ, m,$ and $W$. The convexity of $Q_n(θ)$ in a neighborhood of $Θ_I$ directly follows from Assumption 4.4 (ii). For each $p ∈ S^{d−1}$ and $u ∈ ℝ$, let $θ^* ∈ H(p, Θ_I)$ and $λ^*$ be such that $⟨p, λ^*⟩ ≥ u$ and

$$∥G_{J}(θ^*) + Π_{J}(θ^*)λ^*W_{J}(θ^*)∥ < \inf_{λ ∈ K_{u,p}}∥G_{J}(θ^*) + Π_{J}(θ^*)λW_{J}(θ^*)∥^2 + ε$$

Note that $G_{J}(θ^*)$ only shifts both sides of the inequality in the same manner. Therefore $λ^*$ depends only on the constant matrices $Π_{J}(θ^*)$ and $W_{J}(θ^*)$. For $n$ sufficiently large, by Assumptions 4.4 (i-a) and (i-c), it follows that $ζ_ε(θ^*, λ^*) < ι_{R_{u,p}}ζ_ε(θ, λ) + ε$ with probability one. This ensures the $ε$-argmin condition. For any finite $α$ and $n$ large, the level-$α$ set of $ζ_ε(θ, λ)$ must be a subset of $Θ_I × L_n$ for some bounded set $L_n$ since $ζ_ε(θ, λ) → ∞$ if $θ ∉ Θ_I$.

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 $ζ_ε(θ, λ) = ϕ(ζ_n(θ, λ) + θ, λ)$.

Proof of Corollary 4.2. Let $s : ∂Θ × ℝ^d → ℝ^J(θ)$ be a vector-valued mapping whose $j$-th component is $s_j(θ, λ) = 1\{G_j(θ) + Π_j(θ, λ) > 0\}$. 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 $127$:

$$p = 2μΠ_{J}(θ)(G_{J}(θ) + Π_{J}(θ)λ^*) ∗ s(θ, λ^*)$$

$$t ≥ ||W_{J}^{1/2}(θ)(G_{J}(θ) + Π_{J}(θ)λ^*) ∗ s(θ, λ^*)||^2$$

$$0 ≤ μ$$

$$0 = μ(||W_{J}^{1/2}(θ)(G_{J}(θ) + Π_{J}(θ)λ^*) ∗ s(θ, λ^*)||^2 - t),$$

where $μ$ is the Lagrange multiplier associated with the constraint in Eq. (4.3). By Assumption

\[27\] The constraint is non-differentiable only at finite number of points, and the probability of $G(θ)$ taking these values is 0.
4.5 (ii), the constraint in (4.3) binds, and the conditions above simplify to
\[
p = 2\mu \Pi_{\mathcal{J}(\theta)}(\theta)p W_{\mathcal{J}(\theta)}(\theta)(G_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*) \tag{B.32}
\]
\[
t = ||W_{\mathcal{J}(\theta)}^{1/2}(\theta)(G_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)||^2 \tag{B.33}
\]
\[
\mu > 0. \tag{B.34}
\]

We can solve (B.32) to obtain
\[
(W_{\mathcal{J}(\theta)}^{1/2}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')W_{\mathcal{J}(\theta)}^{1/2}(\theta))^{-1}W_{\mathcal{J}(\theta)}^{1/2}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)p
= 2\mu W_{\mathcal{J}(\theta)}^{1/2}(\theta)(G_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*). \tag{B.35}
\]

Let \( \mathcal{R}(p, \theta) \) be the left hand side of the equation above. Take squared norms both sides to obtain
\[
||\mathcal{R}(p, \theta)||^2 = |2\mu|^2||W_{\mathcal{J}(\theta)}^{1/2}(\theta)(G_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)||^2
= |2\mu|^2t,
\]
where the second equality follows from (B.33). So, we obtain
\[
2\mu = ||\mathcal{R}(p, \theta)||^{-1/2}. \tag{B.36}
\]

Plugging this into (B.35) gives
\[
W_{\mathcal{J}(\theta)}^{1/2}(\theta)(G_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*) = \frac{\mathcal{R}(p, \theta)}{||\mathcal{R}(p, \theta)||} t^{1/2}. \tag{B.37}
\]

Substituting (B.36) and (B.37) into (B.32) yields
\[
p = \Pi_{\mathcal{J}(\theta)}' W_{\mathcal{J}(\theta)}^{1/2} \mathcal{R}(p, \theta). \tag{B.38}
\]

Now, we can use this result to obtain
\[
\mathcal{V}(p, \theta, t) = \langle p, \lambda^* \rangle
= \left< \Pi_{\mathcal{J}(\theta)}' W_{\mathcal{J}(\theta)}^{1/2} \mathcal{R}(p, \theta), \lambda^* \right>
= \left< \mathcal{R}(p, \theta), W_{\mathcal{J}(\theta)}^{1/2} \Pi_{\mathcal{J}(\theta)} \lambda^* \right>
= \left< \mathcal{R}(p, \theta), W_{\mathcal{J}(\theta)}^{1/2}(\Pi_{\mathcal{J}(\theta)} \lambda^* \circ s(\theta, \lambda^*)) \right>
= \left< \mathcal{R}(p, \theta), \frac{\mathcal{R}(p, \theta)}{||\mathcal{R}(p, \theta)||} t^{1/2} - W_{\mathcal{J}(\theta)}^{1/2}(\theta)(G_{\mathcal{J}(\theta)} \circ s(\theta, \lambda^*)) \right>
= ||\mathcal{R}(p, \theta)||^{1/2} - \left< \mathcal{R}(p, \theta), W_{\mathcal{J}(\theta)}^{1/2}(\theta)G_{\mathcal{J}(\theta)} \right>,
\]
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 (B.37).
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. (B.38) 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 \( Z(p, t) := \mu(t) + Z^*(p) \) with \( \mu(t) = t^{1/2} \) and
\[
Z^*(p) = \sup_{\theta \in H(p, \Theta_I)} -\langle \mathcal{R}(p, \theta), W_{\mathcal{J}(\theta)}^{1/2}(\theta) \mathcal{G}_{\mathcal{J}(\theta)}(\theta) \rangle = \sup_{\theta \in H(p, \Theta_I)} -((\Pi_{\mathcal{J}(\theta)}(\Pi_{\mathcal{J}(\theta)})(\theta))^{-1} \Pi_{\mathcal{J}(\theta)}(\theta)p, \mathcal{G}_{\mathcal{J}(\theta)}(\theta)).
\]

For Theorem 4.2, we require the following regularity conditions.

**Assumption B.2 (Local Process Regularity for QLR Statistic):** (i) For any finite sets \( U \subset \mathbb{R} \) and \( S \subset \mathbb{S}^{d-1} \), \( (\sup_{R_u} \tilde{\zeta}_n, (u, p) \in U \times S) \xrightarrow{d} (\sup_{R_u} \tilde{\zeta}, (u, p) \in U \times S) \). (ii) For any \( 0 < \epsilon, \) there exists \( \delta > 0 \) such that
\[
\lim_{n \to \infty} P \left( \sup_{\|p-q\|<\delta} \left| \sup_{R_u} \tilde{\zeta}_n(\theta, \lambda) - \sup_{R_u} \tilde{\zeta}(\theta, \lambda) \right| \geq \epsilon \right) \leq \epsilon,
\]
where \( R_u := H(p, \Theta_I) \times K_u. \)

Assumption B.2 (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 epicconvergence. 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
\[
\mathcal{L}(p, u) := \sup_{\theta \in H(p, \Theta_I)} \sup_{\lambda \in K_u} \| W^{1/2}(\theta)(\mathcal{G}(\theta) + \Pi(\theta)) \|^2. \]
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 \( j \in \mathcal{J}(\theta) \) are irrelevant. To obtain a closed form for \( \mathcal{L} \), consider the following optimization problem
\[
\mathcal{C}(\theta, p, u) := \sup_{\lambda} \| W^{1/2}(\theta)(\mathcal{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda) \|^2 \quad \text{s.t.} (p, \lambda) \leq u. \quad (B.39)
\]
Similar to the proof of Corollary 4.2, the solution \( \lambda^* \) of the problem above satisfies the
following KKT conditions with probability 1.

\[ \nu p = 2\Pi_{\mathcal{J}(\theta)}(\theta)^\prime W_{\mathcal{J}(\theta)}(\theta)\left(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\right) \circ s(\theta, \lambda^*) \]

\[ \langle p, \lambda^* \rangle \leq u \]

\[ 0 \leq \nu \]

\[ 0 = \nu(u - \langle p, \lambda^* \rangle), \]

where \( \nu \) is the Lagrange multiplier associated with the constraint in (B.39). By Assumption 4.5 (ii), the constraint in (B.39) binds, and the conditions above simplify to

\[ \nu p = 2\Pi_{\mathcal{J}(\theta)}(\theta)^\prime W_{\mathcal{J}(\theta)}(\theta)\left(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\right) \circ s(\theta, \lambda^*) \quad \text{(B.40)} \]

\[ \langle p, \lambda^* \rangle = u \quad \text{(B.41)} \]

We can solve (B.40) to obtain

\[ \nu \mathcal{R}(p, \theta) = 2W_{\mathcal{J}(\theta)}(\theta)\left(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\right) \circ s(\theta, \lambda^*). \quad \text{(B.42)} \]

Taking squared norms both sides, we obtain

\[ \nu^2\|\mathcal{R}(p, \theta)\|^2 = 4\|W_{\mathcal{J}(\theta)}(\theta)\left(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\right) \circ s(\theta, \lambda^*)\|^2 \]

\[ = 4\mathcal{C}(\theta, p, u). \quad \text{(B.43)} \]

Plugging in \( \nu = 2\mathcal{C}(\theta, p, u)^{1/2}/\|\mathcal{R}(p, \theta)\| \) back to (B.40), we obtain

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

Now, substitute this into (B.41),

\[ u = \|\mathcal{R}(p, \theta)\|\mathcal{C}(\theta, p, u)^{-1/2}\left\langle \Pi_{\mathcal{J}(\theta)}(\theta)^\prime W_{\mathcal{J}(\theta)}(\theta)\left(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\right) \circ s(\theta, \lambda^*), \lambda^* \right\rangle \]

\[ = \|\mathcal{R}(p, \theta)\|\mathcal{C}(\theta, p, u)^{-1/2}\left\langle W_{\mathcal{J}(\theta)}(\theta)\left(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*\right) \circ s(\theta, \lambda^*), W_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^* \right\rangle \]

\[ = \left\langle \mathcal{R}(p, \theta), W_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^* \right\rangle, \]

where the second equality follows from (B.42). Using (B.42) and the result above, the right hand side of (B.43) can be alternatively written as

\[ 2\nu \left( \left\langle \mathcal{R}(p, \theta), W_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta) \right\rangle + \left\langle \mathcal{R}(p, \theta), W_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^* \right\rangle \right) \]

\[ = 2\nu \left( \left\langle \mathcal{R}(p, \theta), W_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta) \right\rangle + u \right). \]
Therefore, from (B.43), we obtain
\[
\nu = 2\|R(p, \theta)\|^{-1} \left( \left\langle R(p, \theta), W_{J(\theta)}^{1/2}(\theta)G_{\mathcal{J}(\theta)}(\theta) \right\rangle + u \right)
\]
\[
= 2\|R(p, \theta)\|^{-1} \left( \left\langle R(p, \theta), W_{J(\theta)}^{1/2}(\theta)G_{\mathcal{J}(\theta)}(\theta) \right\rangle + u \right)_+ ,
\]
where the second equality follows from the fact \(\nu > 0\). As \(C(\theta, p, u) = \|R(p, \theta)\|^2/4\), we have
\[
C(\theta, p, u) = \|R(p, \theta)\|^{-1} \left( \left\langle R(p, \theta), W_{J(\theta)}^{1/2}(\theta)G_{\mathcal{J}(\theta)}(\theta) \right\rangle + u \right)_+ ^2.
\]
Take the supremum over \(H(p, \Theta_I)\). The result follows.

Proof of Corollary 4.3. We first analyze the Wald statistic \(\sup_{p \in \mathbb{S}^{d-1}} \{-Z_n(p, t) + t^{1/2}\}_+^2\). By Corollary 4.2, the distributional limit \(\sup_{p \in \mathbb{S}^{d-1}} \{-Z(p, t) + t^{1/2}\}_+^2\) of this statistic can be represented as
\[
\sup_{p \in \mathbb{S}^{d-1}} \left\{ \sup_{\theta \in H(p, \Theta_I)} - \left\langle (\Pi_{J(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)^{\prime})^{-1}\Pi_{J(\theta)}(\theta)p, G_{\mathcal{J}(\theta)}(\theta) \right\rangle \right\}^2 
\]
\[
= \sup_{p \in \mathbb{S}^{d-1}} \left\{ \inf_{\theta \in H(p, \Theta_I)} \left\langle (\Pi_{J(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)^{\prime})^{-1}\Pi_{J(\theta)}(\theta)p, G_{\mathcal{J}(\theta)}(\theta) \right\rangle \right\}^2 
\]
\[
= \sup_{p \in \mathbb{S}^{d-1}} \left\langle (\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))^{\prime})^{-1}\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))p, G_{\mathcal{J}(\theta_I(p))}(\theta_I(p)) \right\rangle^2 
\]
\[
= Z,
\]
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
\[
\sup_{p \in \mathbb{S}^{d-1}} \sup_{\theta \in H(p, \Theta_I)} \left\langle R(p, \theta), W_{J(\theta)}^{1/2}(\theta)G_{\mathcal{J}(\theta)}(\theta) \right\rangle^2 
\]
\[
= \sup_{p \in \mathbb{S}^{d-1}} \left\langle (\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))^{\prime})^{-1}\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))p, G_{\mathcal{J}(\theta_I(p))}(\theta_I(p)) \right\rangle^2 = Z.
\]
For the second part, note that \(\tau^*_{1-\alpha}\) is the \(1 - \alpha\) quantile of \(Z\). Therefore, it suffices to show
that $t_{1-\alpha}^*$ is also the $1-\alpha$ quantile of $Z$ under our hypotheses. For that, we can write

$$t_{1-\alpha}^* = \inf \left\{ t : P \left( \sup_{p \in S} \{-Z(p, t)\}_+ \leq 0 \right) \geq 1 - \alpha \right\}$$

$$= \inf \left\{ t : P \left( \sup_{p \in S} \{-t^{1/2} - Z^*(p)\}_+ \leq 0 \right) \geq 1 - \alpha \right\}$$

$$= \inf \left\{ t : P \left( \sup_{p \in S} \{-Z^*(p)\}_+ \leq t^{1/2} \right) \geq 1 - \alpha \right\}$$

$$= \inf \left\{ t : P \left( \sup_{p \in S} \{-Z^*(p)\}_+^2 \leq t \right) \geq 1 - \alpha \right\}$$

$$= \inf \left\{ t : P(Z \leq t) \geq 1 - \alpha \right\},$$

where the third equality follows from the fact that for any $x \geq 0$ and a continuous function $f$, $\sup_{p \in S} \{-x + f(p)\}_+ \leq 0 \Leftrightarrow \sup_{p} \{ f(p) \}_+ \leq x$. \qed

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

**Lemma B.12**: 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 $Q$ be a countable dense subset of $\mathbb{R}^d$. If $\xi_n \overset{f.d.}{\to} \xi$ on $Q$ where $P(\xi \in \Gamma(\mathbb{R}^d)) = 1$, then $\xi_n \overset{e.d.}{\to} \xi$.

**Proof of Lemma B.12.** See Lemma 3.1. in Geyer (2003). \qed

**Proof of Theorem 4.4.** Let

$$\tilde{\zeta}_n(\theta, \lambda) = nQ_n(\theta + \lambda/\sqrt{n})$$

$$= (\sqrt{n}(\tilde{E}_n(X_{1i}) - \theta_1) - \lambda + \sqrt{n}(\theta_1 - \theta))^2_+ + (\sqrt{n}(\theta_2 - \theta_2) + \lambda - \sqrt{n}(\tilde{E}_n(X_{2i}) - \theta_2))^2_+ + \infty \times 1_{\theta \notin \Theta}.$$ 

This function is convex in $(\theta, \lambda)$, 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

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

This function is convex and lsc, and finite on an open interval $(\theta_1, \theta_2)$, and $\tilde{\zeta}_n(\theta, \lambda) \overset{f.d.}{\to} \tilde{\zeta}(\theta, \lambda)$. Therefore, Lemma B.12 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 \( Z \). For example, when \( p = -1 \) and \( \theta \in H(-1, \theta) = \{ \theta_1 \} \), we have \( J(\theta_1) = 1 \), \( R(-1, \theta) = 1 \). Therefore,
\[
Z(-1, t) = t^{1/2} - Z_1.
\]

Similarly,
\[
Z(1, t) = t^{1/2} + Z_2.
\]

Therefore, the limiting process \( Z(p, t) \) has mean \( t^{1/2} \) and covariance \( E[(Z(-1, t) - t^{1/2})(Z(-1, t) - t^{1/2})] = \Omega_{11}, \; E[(Z(1, t) - t^{1/2})(Z(1, t) - t^{1/2})] = \Omega_{22}, \) and \( E[(Z(-1, t) - t^{1/2})(Z(1, t) - t^{1/2})] = -\Omega_{12} \). By Corollary 4.1,
\[
\sqrt{n} d_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \max\{|Z(-1, t)|, |Z(1, t)|\}
\]
and
\[
\sqrt{n} d_H(\Theta_I, \hat{\Theta}_n(t)) \xrightarrow{d} \max\{-Z(-1, t)_+, -Z(1, t)_+\}. \quad \Box
\]

**Proof of Theorem 4.3.** The result for \( \mathcal{W}_n \) follows directly from Theorem 4.4. The results for \( \mathcal{QLR}_n \) and \( \tilde{\mathcal{W}}_n \) are due to CHT and BM respectively. \quad \Box