Journal of Econometrics 192 (2016) 269–290

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A dual approach to inference for partially identified econometric models

Hiroaki Kaido

Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA

ARTICLE INFO

Article history:
Received 27 March 2012
Received in revised form
23 September 2015
Accepted 21 December 2015
Available online 8 January 2016

JEL classification:
C12

Keywords:
Partial identification
Criterion function
Support function

ABSTRACT

This paper considers inference for the set Θ_I of parameter values that minimize a criterion function. Chernozhukov et al. (2007) (CHT) develop a general theory of estimation and inference using the level-set of a criterion function. We establish a dual relationship between the level-set estimator and its support function and show that the properly normalized support function provides alternative Wald-type inference methods. These methods can be used to obtain confidence sets for Θ_I and points inside it. For models with finitely many moment inequalities, we show that our Wald-type statistic is asymptotically equivalent to CHT's statistic under regularity conditions.

1. Introduction

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

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

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

In this paper, we focus on econometric models with compact convex identified sets, which enables us to characterize the identified set by its support function.[1] This class includes many econometric models studied recently, e.g., regression with interval data (Manski and Tamer, 2002; Magnac and Maurin, 2008), a class of discrete choice models (Pakes, 2010), consumer demand models with unobserved heterogeneity (Blundell et al., 2014), and an asset pricing model in incomplete markets (Kaido and White, 2009). Following CHT, our estimator of Θ_I is the level set $\hat{\Theta}_n = \{\theta$:

E-mail address: hkaido@bu.edu.

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

http://dx.doi.org/10.1016/j.jeconom.2015.12.017
0304-4076/

$Q_n(\theta) \le t_n\}$ of a criterion function $Q_n(\cdot)$ for some sequence of levels $\{t_n\}$. The support function approach provides a straightforward algorithm to compute the boundary of this estimator. Specifically, we propose to solve the optimization problem $\max_{Q_n(\theta)\le t_n}\langle p,\theta\rangle$ for each p. This yields the support function $s(\cdot,\hat{\Theta}_n)$ of the set estimator as a value function and also gives the boundary of $\hat{\Theta}_n$. The optimization is a convex programming problem, which can be solved using standard algorithms.

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

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

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

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

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

2. General setup

2.1. Criterion functions and set estimator

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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