Title
Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities

Journal
Econometrica, 82(1)

ISSN
0012-9682

Authors
Kaido, Hiroaki
Santos, Andres

Publication Date
2014

DOI
10.3982/ecta10017

Peer reviewed

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 82, No. 1 (January, 2014), 387–413

ASYMPTOTICALLY EFFICIENT ESTIMATION OF MODELS DEFINED BY CONVEX MOMENT INEQUALITIES

HIROAKI KAIDO
Boston University, Boston, MA 02215, U.S.A.

ANDRES SANTOS
University of California—San Diego, La Jolla, CA 92093-0508, U.S.A.

Econometrica, Vol. 82, No. 1 (January, 2014), 387–413

NOTES AND COMMENTS

ASYMPTOTICALLY EFFICIENT ESTIMATION OF MODELS DEFINED BY CONVEX MOMENT INEQUALITIES

BY HIROAKI KAIDO AND ANDRES SANTOS[1]

This paper examines the efficient estimation of partially identified models defined by moment inequalities that are convex in the parameter of interest. In such a setting, the identified set is itself convex and hence fully characterized by its support function. We provide conditions under which, despite being an infinite dimensional parameter, the support function admits $\sqrt{n}$-consistent regular estimators. A semiparametric efficiency bound is then derived for its estimation, and it is shown that any regular estimator attaining it must also minimize a wide class of asymptotic loss functions. In addition, we show that the "plug-in" estimator is efficient, and devise a consistent bootstrap procedure for estimating its limiting distribution. The setting we examine is related to an incomplete linear model studied in Beresteanu and Molinari (2008) and Bontemps, Magnac, and Maurin (2012), which further enables us to establish the semiparametric efficiency of their proposed estimators for that problem.

KEYWORDS: Semiparametric efficiency, partial identification, moment inequalities.

1. INTRODUCTION

IN A LARGE NUMBER OF ESTIMATION PROBLEMS, the data available to the researcher fail to point identify the parameter of interest, but are still able to bound it in a potentially informative way (Manski (2003)). This phenomenon has been shown to be common in economics, where partial identification arises naturally as the result of equilibrium behavior in game theoretic contexts (Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2011)), certain forms of censoring (Manski and Tamer (2002)), and optimal behavior by agents in discrete choice problems (Pakes, Porter, Ho, and Ishii (2006), Pakes (2010)).

A common feature of many of these settings is that the bounds on the parameter of interest are implicitly determined by moment inequalities. Specifically, let $X_i \in \mathcal{X} \subseteq \mathbf{R}^{d_X}$ be a random vector with distribution P, $\Theta \subset \mathbf{R}^{d_\theta}$ denote the parameter space, and $m : \mathcal{X} \times \Theta \to \mathbf{R}^{d_m}$ and $F : \mathbf{R}^{d_m} \to \mathbf{R}^{d_F}$ be known functions. In many models, the identified set is of the general form

$$\Theta_0(P) \equiv \left\{ \theta \in \Theta : F\left(\int m(x, \theta)\, dP(x) \right) \le 0 \right\}. \tag{1}$$

[1]We would like to thank Francesca Molinari, James Stock, and two anonymous referees for valuable suggestions that helped greatly improve this paper. We are also indebted to Victor Chernozhukov, Mark Machina, Ulrich Müller, and numerous seminar participants for their helpful comments.

 DOI: 10.3982/ECTA10017

A prevalent specification is one in which F is the identity mapping, in which case (1) reduces to the moment inequalities model studied in Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2010), and Andrews and Soares (2010), among others. Examples where F is not the identity include binary choice models with misclassified or endogenous regressors (Mahajan (2003), Chesher (2009)).

We contribute to the existing literature by developing an asymptotic efficiency concept for estimating an important subset of these models. Heuristically, estimation of the identified set is tantamount to estimation of its boundary. In obtaining an asymptotic efficiency result, it is therefore instrumental to characterize the boundary of the identified set as a function of the unknown distribution P. We obtain such a characterization in the special, yet widely applicable, setting in which the constraint functions are convex, for example linear, in θ. In such instances, the identified set is itself convex and its boundary is determined by the hyperplanes that are tangent to it. The set of tangent, or supporting, hyperplanes can, in turn, be identified with a unique function on the unit sphere called the *support function* of the identified set. As a result, estimation of the identified set may be accomplished through the estimation of its support function—an insight previously exploited by Beresteanu and Molinari (2008), Bontemps, Magnac, and Maurin (2012), and Kaido (2012).

We provide conditions under which, despite being an infinite dimensional parameter, the support function of the identified set admits $\sqrt{n}$-consistent regular estimators. By way of the convolution theorem, we further establish that any regular estimator of the support function must converge in distribution to the sum of an "efficient" mean zero Gaussian process $\mathbb{G}_0$ and an independent "noise" process Δ_0. In accord with finite dimensional problems, an estimator is therefore considered to be semiparametrically efficient if it is regular and its asymptotic distribution equals that of $\mathbb{G}_0$; that is, its corresponding noise process Δ_0 equals zero almost surely. Obtaining a semiparametric efficiency bound then amounts to characterizing the distribution of $\mathbb{G}_0$, which in finite dimensional problems is equivalent to reporting its covariance matrix. In the present context, we obtain the semiparametric efficiency bound by deriving the covariance kernel of the Gaussian process $\mathbb{G}_0$. These insights are readily applicable to other convex partially identified models, a point we illustrate by showing that the estimators proposed in Beresteanu and Molinari (2008) and Bontemps, Magnac, and Maurin (2012) are efficient.

Among the implications of semiparametric efficiency is that an efficient estimator minimizes diverse measures of asymptotic risk among regular estimators. Due to the close link between convex sets and their support functions, optimality in estimating the support function of the identified set further leads to optimality in estimating the identified set itself. Specifically, we show that, among regular convex set estimators, the set associated with the efficient estimator for the support function minimizes asymptotic risk for a wide class of loss functions based on Hausdorff distance. These results complement Song

(2012), who employed semiparametric efficient estimators of interval valued identified sets to construct an optimal statistic for researchers that must make a point decision.

Having characterized the semiparametric efficiency bound, we establish that the support function of the sample analog to (1) is the efficient estimator. A consequence of this result is that the sample analog is also efficient for estimating the "marginal" identified set of any particular coordinate of the vector θ. Interestingly, regular estimation of the support functions of these marginal identified sets requires weaker assumptions than those needed to obtain a regular estimator of the support function of $\Theta_0(P)$. Finally, we conclude by constructing a bootstrap procedure for consistently estimating the distribution of the efficient limiting process $\mathbb{G}_0$. We illustrate the applicability of this result by constructing inferential procedures that are pointwise (in P) consistent in level.

In related work, Beresteanu and Molinari (2008) first employed support functions in the study of partially identified models. The authors derived methods for conducting inference on the identified set through its support function, providing insights we rely upon in our analysis. The use of support functions to characterize semiparametric efficiency, however, is novel to this paper. Other work on estimation includes Hirano and Porter (2012), who obtained conditions under which no regular estimators exist in intersection bounds models for scalar valued parameters, and Song (2010), who proposed robust estimators for such problems. Our results complement theirs by clarifying what the sources of irregularity are in settings where the parameter of interest has dimension greater than 1.

A large literature on the moment inequalities model has focused on the complementary problem of inference. The framework we employ is not as general as the one pursued in these papers, which, for example, do not impose convexity; see Romano and Shaikh (2008), Andrews and Guggenberger (2009), Rosen (2008), Menzel (2009), Bugni (2010), Canay (2010), and Andrews and Barwick (2012), among others. This paper is also part of the literature on efficient estimation in econometrics, which has primarily studied finite dimensional parameters identified by moment equality restrictions; see Chamberlain (1987, 1992), Brown and Newey (1998), Ai and Chen (2009), and references therein.

The remainder of the paper is organized as follows. Section 2 introduces the moment inequalities we study and examples of models that fall within its scope. In Section 3, we characterize the efficiency bound, while in Section 4, we show that the plug-in estimator is efficient. Section 5 derives the consistent bootstrap procedure. The Supplemental Material (Kaido and Santos (2014)) contains all proofs and a Monte Carlo study.

2. GENERAL SETUP

It will prove helpful to consider the identified set as a function of the unknown distribution of X_i. For this reason, we make such dependence explicit by defining the *identified set under Q* to be

$$\Theta_0(Q) \equiv \left\{\theta \in \Theta : F\left(\int m(x, \theta)\, dQ(x)\right) \leq 0\right\}.$$

Thus, $\Theta_0(Q)$ is the set of parameter values that is identified by the moment restrictions when data are generated according to the probability measure Q. We may then interpret the actual identified set $\Theta_0(P)$ as the value the known mapping $Q \mapsto \Theta_0(Q)$ takes at the unknown distribution P.

Our analysis focuses on settings where the identified set is convex, which we ensure by requiring that the functions $\theta \mapsto F^{(i)}(\int m(x, \theta)\, dP(x))$ be themselves convex for all $1 \leq i \leq d_F$—here and throughout, $w^{(i)}$ denotes the ith coordinate of a vector w. Unfortunately, convexity is not sufficient for establishing that $\Theta_0(P)$ admits a regular estimator. In particular, special care must be taken when a constraint function is linear in θ leading to a “flat face” in the boundary of the identified set. We will show by example that when the slope of a linear constraint depends on the underlying distribution, a small perturbation of P may lead to a nondifferentiable change in the identified set. This lack of differentiability in turn implies that there exist no asymptotically linear regular estimators (van der Vaart (1991), Hirano and Porter (2012)).

For this reason, we assume that the slope of any linear constraint is known. Specifically, we let

$$m(x, \theta) \equiv \left(m_S(x, \theta)', \theta' A'\right)', \tag{2}$$

where $m_S : \mathcal{X} \times \Theta \to \mathbf{R}^{d_{m_S}}$ is a known measurable function, and A is a known $d_F \times d_\theta$ matrix. For an also known function $F_S : \mathbf{R}^{d_{m_S}} \to \mathbf{R}^{d_F}$, we then assume $F : \mathbf{R}^{d_m} \to \mathbf{R}^{d_F}$ satisfies

$$F\left(\int m(x, \theta)\, dP(x)\right) = A\theta + F_S\left(\int m_S(x, \theta)\, dP(x)\right), \tag{3}$$

where, for each $1 \leq i \leq d_F$, the function $\theta \mapsto F_S^{(i)}(\int m_S(x, \theta)\, dP(x))$ may only depend on a subvector of θ, but is required to be strictly convex in this subvector. Formally, let $\mathcal{S}_i \subseteq \{1, \ldots, d_\theta\}$ denote the smallest set such that, if $\theta_1, \theta_2 \in \Theta$ satisfy $\theta_1^{(j)} = \theta_2^{(j)}$ for all $j \in \mathcal{S}_i$, then

$$F_S^{(i)}\left(\int m_S(x, \theta_1)\, dQ(x)\right) = F_S^{(i)}\left(\int m_S(x, \theta_2)\, dQ(x)\right) \tag{4}$$

for all Borel measures Q on $\mathcal{X}$.[2] We then refer to the *arguments* of $\theta \mapsto F_S^{(i)}(\int m_S(x,\theta)\,dP(x))$ as the coordinates of θ corresponding to indices in $\mathcal{S}_i$, and require that, for all $\lambda \in (0,1)$,

$$F_S^{(i)}\left(\int m_S\big(x, \lambda\theta_1 + (1-\lambda)\theta_2\big)\,dP(x)\right) < \lambda F_S^{(i)}\left(\int m_S(x,\theta_1)\,dP(x)\right) + (1-\lambda)F_S^{(i)}\left(\int m_S(x,\theta_2)\,dP(x)\right)$$

whenever $\theta_1^{(j)} \neq \theta_2^{(j)}$ for some $j \in \mathcal{S}_i$. For instance, if $\mathcal{S}_i = \emptyset$, then, by (3), constraint i is linear in θ with known slope but intercept potentially depending on P. Similarly, if $\mathcal{S}_i = \{1, \dots, d_\theta\}$, then constraint i is strictly convex in θ. In between these cases are specifications of the constraints that are linear in some parameters and strictly convex in others.

As a final piece of notation, it will prove helpful to index the constraints that are active at each point θ in an identified set $\Theta_0(Q)$. Toward this end, for each $\theta \in \Theta_0(Q)$, we define

$$\mathcal{A}(\theta, Q) \equiv \left\{ i \in \{1, \dots, d_F\} : F^{(i)}\left(\int m(x,\theta)\,dQ(x)\right) = 0 \right\}.$$

2.1. *Examples*

In order to fix ideas, we briefly discuss applications of our general framework. We revisit these examples in more detail in the Supplemental Material, where we additionally examine the implications of our regularity conditions and provide sufficient conditions for them to hold.

Our first example is a special case of the analysis in Manski and Tamer (2002).

EXAMPLE 2.1—Interval Censored Outcome: An outcome variable Y is generated according to

$$Y = Z'\theta_0 + \varepsilon,$$

where $Z \in \mathbf{R}^{d_\theta}$ is a regressor with discrete support $\mathcal{Z} \equiv \{z_1, \dots, z_K\}$ and ε satisfies $E[\varepsilon|Z] = 0$. Suppose Y is unobservable, but there exist (Y_L, Y_U) such that $Y_L \le Y \le Y_U$ almost surely. The identified set for θ_0 then consists of all parameters $\theta \in \Theta$ satisfying the inequalities

$$E[Y_L|Z = z_k] - z_k'\theta \le 0, \quad k = 1, \dots, K,$$
$$z_k'\theta - E[Y_U|Z = z_k] \le 0, \quad k = 1, \dots, K.$$

[2]If $A \subseteq \{1, \dots, d_\theta\}$ and $B \subseteq \{1, \dots, d_\theta\}$ satisfy (4), then so does $A \cap B$, implying $\mathcal{S}_i$ is well defined.

These inequalities can be written as in (3) with $F_S^{(i)}(\int m_S(x,\theta)\,dP(x))$ equal to $E[Y_L|Z=z_k]$ or $-E[Y_U|Z=z_k]$ for some k. Note that all constraints are linear, and hence $\mathcal{S}_i=\emptyset$ for all i.

Another prominent application of moment inequality models is in the context of discrete choice.

EXAMPLE 2.2—Discrete Choice: Suppose an agent chooses $z\in\mathbf{R}^{d_Z}$ from a set $\mathcal{Z}\equiv\{z_1,\dots,z_K\}$ so as to maximize his expected payoff $E[\pi(Y,Z,\theta_0)|\mathcal{F}]$, where Y is a vector of observable random variables and $\mathcal{F}$ is the agent's information set. Letting $z^*\in\mathcal{Z}$ denote the optimal choice, we obtain

$$E\big[\pi(Y,z,\theta_0)-\pi\big(Y,z^*,\theta_0\big)|\mathcal{F}\big]\le 0 \tag{5}$$

for all $z\in\mathcal{Z}$. A common specification is that $\pi(y,z,\theta_0)=\psi(y,z)+z'\theta_0$; see Pakes et al. (2006) and Pakes (2010). Therefore, under suitable assumptions on the agent's beliefs, the optimality conditions in (5) then imply that θ_0 must satisfy the moment inequalities

$$E\big[\big(\big(\psi(Y,z_j)-\psi(Y,z_k)\big)+(z_j-z_k)'\theta_0\big)1\big\{Z^*=z_k\big\}\big]\le 0 \tag{6}$$

for any $z_j,z_k\in\mathcal{Z}$. As in Example 2.1, the restrictions in (6) may be expressed in the form of (3).

Strictly convex moment inequalities arise in asset pricing (Hansen, Heaton, and Luttmer (1995)).[3]

EXAMPLE 2.3—Pricing Kernel: Let $Z\in\mathbf{R}^{d_Z}$ denote the payoffs of d_Z securities which are traded at a price of $U\in\mathbf{R}^{d_Z}$. If short sales are not allowed for any securities, then the feasible set of portfolio weights is restricted to $\mathbf{R}_+^{d_Z}$ and the standard Euler equation does not hold. Instead, under power utility, Luttmer (1996) derived a modified (unconditional) Euler equation of the form

$$E\left[\frac{1}{1+\rho}Y^{-\gamma}Z-U\right]\le 0, \tag{7}$$

where Y is the ratio of future over present consumption, ρ is the investor's subjective discount rate, and γ is the relative risk aversion coefficient. If $Z^{(i)}\ge 0$ almost surely and $Z^{(i)}>0$ with positive probability, then the constraints in (7) are strictly convex in $\theta=(\rho,\gamma)'\in\mathbf{R}^2$. To map (7) into (3), we let $A=0$ and $F_S^{(i)}(\int m_S(x,\theta)\,dP(x))=E[\frac{1}{1+\rho}Y^{-\gamma}Z^{(i)}-U^{(i)}]$, implying $\mathcal{S}_i=\{1,2\}$ for all i.

[3]We note our semiparametric efficiency bound is for i.i.d. data and requires an extension to time series for its applicability to asset pricing. Example 2.3 is nonetheless introduced to illustrate the role of strictly convex constraints.

The following example is based on the discussion in Blundell and MaCurdy (1999).

EXAMPLE 2.4—Participation Constraint: Consider an agent with Stone–Geary preferences over consumption $C \in \mathbf{R}_+$ and leisure $L \in [0, T]$ parameterized by

$$u(C, L) = \log(C - \alpha) + \beta \log(L).$$

Given wage W and non-labor income $Y \in \mathbf{R}_+$, the agent maximizes expected utility subject to the budget constraint $C = Y + W(T - L)$ and the constraint $0 \leq L \leq T$. If Y is unknown to the agent when the labor decision is made, then her first order conditions imply

$$E\left[\left(\frac{W}{C-\alpha} - \frac{\beta}{L}\right)Z\right] = E\left[E\left[\frac{W}{C-\alpha} - \frac{\beta}{L}|\mathcal{F}\right]Z\right] \leq 0, \tag{8}$$

where $\mathcal{F}$ is the information available to the agent when choosing L, and Z is any positive $\mathcal{F}$-measurable random vector. For $\theta = (\alpha, \beta)'$, in this example $\mathcal{S}_i = \{1\}$ for all i.

3. SEMIPARAMETRIC EFFICIENCY

3.1. *Preliminaries*

Throughout, we let $\langle p, q\rangle = p'q$ denote the Euclidean inner product of two vectors $p, q \in \mathbf{R}^{d_\theta}$ and $\|p\| = \langle p, p\rangle^{1/2}$ be the Euclidean norm. Following the literature, we employ the Hausdorff metric to evaluate distance between sets in $\mathbf{R}^{d_\theta}$. Hence, for any closed sets A and B, we let

$$d_H(A, B) \equiv \max\{\vec{d}_H(A, B), \vec{d}_H(B, A)\},$$
$$\vec{d}_H(A, B) \equiv \sup_{a\in A} \inf_{b\in B} \|a - b\|,$$

where d_H and $\vec{d}_H$ are the Hausdorff and directed Hausdorff distances, respectively.

For $\mathbb{S}^{d_\theta} \equiv \{p \in \mathbf{R}^{d_\theta} : \|p\| = 1\}$ the unit sphere in $\mathbf{R}^{d_\theta}$, we denote by $\mathcal{C}(\mathbb{S}^{d_\theta})$ the space of bounded continuous functions on $\mathbb{S}^{d_\theta}$ and equip $\mathcal{C}(\mathbb{S}^{d_\theta})$ with the supremum norm $\|f\|_\infty \equiv \sup_{p\in\mathbb{S}^{d_\theta}} |f(p)|$. The support function $\nu(\cdot, K) : \mathbb{S}^{d_\theta} \to \mathbf{R}$ of a compact convex set $K \subset \mathbf{R}^{d_\theta}$ is then given by

$$\nu(p, K) \equiv \sup_{k\in K} \langle p, k\rangle, \quad p \in \mathbb{S}^{d_\theta}.$$

Heuristically, the support function assigns to each vector p the signed distance between the origin and the hyperplane orthogonal to p that is tangent to K. By

Hörmander's embedding theorem, the support functions of any two compact convex sets K_1 and K_2 belong to $\mathcal{C}(\mathbb{S}^{d_\theta})$, and in addition,

$$d_H(K_1, K_2) = \sup_{p \in \mathbb{S}^{d_\theta}} \left| \nu(p, K_1) - \nu(p, K_2) \right|. \tag{9}$$

Therefore, convex compact sets can be identified in a precise sense with elements of $\mathcal{C}(\mathbb{S}^{d_\theta})$ in a way that preserves distances; that is, there exists an isometry between them.

In our analysis, we study the identified set $\Theta_0(P)$ which we characterize by its support function

$$\nu\big(p, \Theta_0(P)\big) = \sup_{\theta \in \Theta_0(P)} \langle p, \theta \rangle. \tag{10}$$

As P is unknown, we view $\nu(\cdot, \Theta_0(P))$ as an infinite dimensional parameter defined on $\mathcal{C}(\mathbb{S}^{d_\theta})$ and aim to characterize the semiparametric efficiency bound for its estimation.

3.1.1. *Efficiency in* $\mathcal{C}(\mathbb{S}^{d_\theta})$

We briefly review the concept of semiparametric efficiency as applied to regular infinite dimensional parameters defined on $\mathcal{C}(\mathbb{S}^{d_\theta})$; please refer to Chapter 5 in Bickel, Klassen, Ritov, and Wellner (1993) for a full discussion. Our analysis is done under the assumption that the data are independent and identically distributed (i.i.d.), and hence we start by imposing the following.

ASSUMPTION 3.1: $\{X_i\}_{i=1}^n$ *is an i.i.d. sample with each* X_i *distributed according to* P.

We let $\mathbf{M}$ denote the set of Borel probability measures on $\mathcal{X}$, endowed with the τ-topology,[4] and μ be a positive σ-finite measure such that P is absolutely continuous with respect to μ (denoted $P \ll \mu$). Of particular interest is the set $\mathbf{M}_\mu \equiv \{P \in \mathbf{M} : P \ll \mu\}$, which may be embedded in

$$L^2_\mu \equiv \{f : \mathcal{X} \to \mathbf{R} : \|f\|_{L^2_\mu} < \infty\}, \quad \|f\|^2_{L^2_\mu} \equiv \int f^2(x)\, d\mu(x),$$

via the mapping $Q \mapsto \sqrt{dQ/d\mu}$. A model $\mathbf{P} \subseteq \mathbf{M}_\mu$ is then a collection of probability measures, which can be identified with a subset $\mathbf{S}$ of L^2_μ that is given by

$$\mathbf{S} \equiv \big\{h \in L^2_\mu : h = \sqrt{dQ/d\mu} \text{ for some } Q \in \mathbf{P}\big\}. \tag{11}$$

[4]The τ-topology is the coarsest topology on $\mathbf{M}$ under which the mappings $Q \mapsto \int f(x)\, dQ(x)$ are continuous for all measurable and bounded functions $f : \mathcal{X} \to \mathbf{R}$. Note that unlike the weak topology, continuity of f is not required.

Employing the introduced notation, we then define *curves* and *tangent sets* in the usual manner.

DEFINITION 3.1: A function $h: N \to L^2_\mu$ is a curve in L^2_μ if $N \subseteq \mathbf{R}$ is a neighborhood of zero and $\eta \mapsto h(\eta)$ is continuously Fréchet differentiable on N. For notational simplicity, we write h_η for $h(\eta)$ and let $\dot{h}_\eta$ denote its Fréchet derivative at any point $\eta \in N$.

DEFINITION 3.2: For $\mathbf{S} \subseteq L^2_\mu$ and a function $s \in \mathbf{S}$, the tangent set of $\mathbf{S}$ at s is defined as

$$\dot{\mathbf{S}}^0 \equiv \{\dot{h}_0 : h_\eta \text{ is a curve in } L^2_\mu \text{ with } h_0 = s \text{ and } h_\eta \in \mathbf{S} \text{ for all } \eta\}.$$

The tangent space of $\mathbf{S}$ at s, denoted by $\dot{\mathbf{S}}$, is the closure of the linear span of $\dot{\mathbf{S}}^0$ (in L^2_μ).

Each curve $\eta \mapsto h_\eta$, with $h_\eta \in \mathbf{S}$, can be associated with a quadratic mean differentiable submodel $\eta \mapsto P_\eta \in \mathbf{P}$ by the relation $h_\eta = \sqrt{dP_\eta/d\mu}$. The main difference between the efficiency analysis of finite and infinite dimensional parameters is in the appropriate notion of differentiability. Formally, a parameter defined on $\mathcal{C}(\mathbb{S}^{d_\theta})$ is a mapping $\rho: \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ that assigns to each $Q \in \mathbf{P}$ a function in $\mathcal{C}(\mathbb{S}^{d_\theta})$. In our context, ρ assigns to Q the support function of its identified set; that is, $\rho(Q) = \nu(\cdot, \Theta_0(Q))$. To derive a semiparametric efficiency bound for estimating $\rho(P)$, we require $\rho: \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ to be smooth in the sense of being pathwise weak-differentiable at P.

DEFINITION 3.3: For a model $\mathbf{P} \subseteq \mathbf{M}_\mu$ and a parameter $\rho: \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$, we say ρ is pathwise weak-differentiable at P if there is a continuous linear operator $\dot{\rho}: \dot{\mathbf{S}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ such that

$$\lim_{\eta \to 0} \left| \int_{\mathbb{S}^{d_\theta}} \left\{ \frac{\rho(h_\eta)(p) - \rho(h_0)(p)}{\eta} - \dot{\rho}(\dot{h}_0)(p) \right\} dB(p) \right| = 0,$$

for any finite Borel measure B on $\mathbb{S}^{d_\theta}$ and any curve $\eta \mapsto h_\eta$ with $h_\eta \in \mathbf{S}$ and $h_0 = \sqrt{dP/d\mu}$.

Given these definitions, we can state a precise notion of semiparametric efficiency for estimating $\rho(P)$ in terms of the convolution theorem. We refer the reader to Theorem 5.2.1 in Bickel et al. (1993) for a more general statement of the convolution theorem and a proof of this result.

THEOREM 3.1—Convolution Theorem: *Suppose*: (i) *Assumption* 3.1 *holds*, (ii) $P \in \mathbf{P}$, (iii) $\dot{\mathbf{S}}^0$ *is linear, and* (iv) $\rho: \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ *is pathwise weak-differentiable* *at P. Then, there exists a tight mean zero Gaussian process $\mathbb{G}_0$ in $\mathcal{C}(\mathbb{S}^{d_\theta})$ such that any regular estimator $\{T_n\}$ of $\rho(P)$ must satisfy*

$$\sqrt{n}\big(T_n - \rho(P)\big) \xrightarrow{L} \mathbb{G}_0 + \Delta_0,$$

where $\xrightarrow{L}$ denotes convergence in law, and Δ_0 is some tight random element independent of $\mathbb{G}_0$.[5]

In complete accord with the finite dimensional setting, the asymptotic distribution of any regular estimator can be characterized as that of a Gaussian process $\mathbb{G}_0$ plus an independent term Δ_0. Thus, a regular estimator may be considered efficient if its asymptotic distribution equals that of $\mathbb{G}_0$. Heuristically, the asymptotic distribution of any competing regular estimator must then equal that of the efficient estimator plus an independent "noise" term. Computing a semiparametric efficiency bound is then equivalent to characterizing the distribution of $\mathbb{G}_0$. In finite dimensional problems, this amounts to computing the covariance matrix of the distributional limit. In the present context, we instead aim to obtain the covariance kernel for the Gaussian process $\mathbb{G}_0$, denoted

$$I^{-1}(p_1, p_2) \equiv \operatorname{Cov}\big(\mathbb{G}_0(p_1), \mathbb{G}_0(p_2)\big),$$

and usually termed the *inverse information covariance functional* for ρ in the model $\mathbf{P}$.

REMARK 3.1: More generally, if a possibly nonconvex identified set $\Theta_0(P)$ is an element of a metric space $\mathbf{B}_1$, then we can consider estimation of the parameter $\rho_1 : \mathbf{P} \to \mathbf{B}_1$ given by $\rho_1(P) = \Theta_0(P)$. However, a key complication in this approach is that $\mathbf{B}_1$ is often not a vector space—a crucial requirement in the theory of semiparametric efficiency. For this reason, in our setting we instead employ an isometry $\rho_2 : \mathbf{B}_1 \to \mathbf{B}_2$ into a Banach space $\mathbf{B}_2$, and examine estimation of $\rho(P) \equiv \rho_2 \circ \rho_1(P)$.[6] This insight is applicable to other partially identified models; for example, a bounded set K can be embedded in L^1_μ through its indicator function. Establishing pathwise weak-differentiability in these contexts, however, will require substantially different arguments than ours.

[5] $\{T_n\}$ is regular if there is a tight Borel measurable $\mathbb{G}$ on $\mathcal{C}(\mathbb{S}^{d_\theta})$ such that, for every curve $\eta \mapsto h_\eta$ in $\mathbf{S}$ passing through $s \equiv \sqrt{dP/d\mu}$ and every $\{\eta_n\}$ with $\eta_n = O(n^{-1/2})$, $\sqrt{n}(T_n - \rho(h_{\eta_n})) \xrightarrow{L_n} \mathbb{G}$, where L_n is the law under $P^n_{\eta_n}$.

[6] Concretely, in our framework, $\mathbf{B}_1$ corresponds to the space of convex compact sets endowed with the Hausdorff metric, $\mathbf{B}_2 = \mathcal{C}(\mathbb{S}^{d_\theta})$ and $\rho_2(K) = \nu(\cdot, K)$ for any $K \in \mathbf{B}_1$.

3.2. *Efficiency Bound*

3.2.1. *Assumptions*

We require the following assumptions to derive the distribution of the efficient limiting process $\mathbb{G}_0$.

ASSUMPTION 3.2: *$\Theta \subset \mathbf{R}^{d_\theta}$ is convex, compact, and has nonempty interior Θ^o (relative to $\mathbf{R}^{d_\theta}$).*

ASSUMPTION 3.3: *The functions $m : \mathcal{X} \times \Theta \to \mathbf{R}^{d_m}$ and $F : \mathbf{R}^{d_m} \to \mathbf{R}^{d_F}$ satisfy (2) and (3).*

ASSUMPTION 3.4: (i) *$m : \mathcal{X} \times \Theta \to \mathbf{R}^{d_m}$ is bounded*; (ii) *$\theta \mapsto m(x, \theta)$ is differentiable at all $x \in \mathcal{X}$ with $\nabla_\theta m(x, \theta)$ bounded in $(x, \theta) \in \mathcal{X} \times \Theta$*; (iii) *$\theta \mapsto \nabla_\theta m(x, \theta)$ is equicontinuous in $x \in \mathcal{X}$.*

ASSUMPTION 3.5: *There is an open set $V_0 \subseteq \mathbf{R}^{d_m}$ such that* (i) *$v \mapsto F(v)$ is differentiable on V_0, and* (ii) *$v \mapsto \nabla F(v)$ is uniformly continuous and bounded on V_0.*

The convexity of Θ can be relaxed provided $m(x, \cdot)$ is well defined on the convex hull of Θ for all $x \in \mathcal{X}$. Assumption 3.4 requires $m(x, \theta)$ and $\nabla_\theta m(x, \theta)$ to be bounded on $\mathcal{X} \times \Theta$, which for some parameterizations implies $\mathcal{X}$ is compact. Assumption 3.5 imposes similar requirements on F.

In addition to Assumptions 3.1–3.5, we need to impose the following requirements on P.

ASSUMPTION 3.6: (i) *$\Theta_0(P) \neq \emptyset$ and $\Theta_0(P) \subset \Theta^o$*; (ii) *there is a neighborhood $N(P) \subseteq \mathbf{M}$ such that, for all $Q \in N(P)$ and $1 \leq i \leq d_F$, the function $\theta \mapsto F_S^{(i)}(\int m_S(x, \theta)\, dQ(x))$ is strictly convex in its arguments*; (iii) *$\int m(x, \theta)\, dP(x) \in V_0$ for all $\theta \in \Theta$*; (iv) *for all $\theta \in \Theta_0(P)$, the vectors $\{\nabla F^{(i)}(\int m(x, \theta)\, dP(x)) \times \int \nabla_\theta m(x, \theta)\, dP(x)\}_{i \in \mathcal{A}(\theta, P)}$ are linearly independent.*

Assumption 3.6(i) implies $\Theta_0(P)$ is characterized by the inequality constraints and not by the parameter space. Certain parameter constraints, however, may be imposed through the moment restrictions; see Remark 3.4. In Assumption 3.6(ii), convexity of the constraints is required at all Q near P (in the τ-topology), which implies $\Theta_0(Q)$ is also convex. Assumption 3.6(iii), together with Assumptions 3.4(ii) and 3.5(ii), ensure the constraints are differentiable in θ. Finally, Assumption 3.6(iv) is the key requirement ensuring $\nu(\cdot, \Theta_0(P))$ is a regular parameter at P. This assumption implies $\Theta_0(P)$ has a nonempty interior, which rules out identification but also ensures that the convex programming problem in (10) satisfies a Slater constraint qualification. The latter result provides us with a sufficient, but not necessary, condition for establishing that $\nu(\cdot, \Theta_0(P))$ has a Lagrangian representation, that is, the duality gap is zero. Additionally, Assumption 3.6(iv) rules out moment equalities, though we note that strictly convex moment equalities would imply either that the model is identified or that the identified set is nonconvex. Interestingly, a violation of Assumption 3.6(iv) is also the condition under which Hirano and Porter (2012) showed irregularity in the problem they studied.

Finally, we define the model $\mathbf{P} \subset \mathbf{M}$ to be the set of probability measures that are dominated by common measure μ, and in addition satisfy Assumption 3.6,

$$\mathbf{P} \equiv \{P \in \mathbf{M} : P \ll \mu \text{ and Assumptions 3.6(i)–(iv) hold}\}.$$

REMARK 3.2: Requiring the slope of linear constraints to be known is demanding but, as we now show, crucial for the support function to be pathwise weak-differentiable. Let $\mathcal{X} \subset \mathbf{R}^2$ be compact, $\Theta \equiv \{\theta \in \mathbf{R}^2 : \|\theta\| \leq B\}$, and denote $x = (x^{(1)}, x^{(2)})'$, $\theta = (\theta^{(1)}, \theta^{(2)})'$. Suppose that, in (1), $F : \mathbf{R}^3 \to \mathbf{R}^3$ is the identity, and that, for some $K > 0$, the function $m : \mathcal{X} \times \Theta \to \mathbf{R}^3$ is given by

$$\begin{aligned} &m^{(1)}(x, \theta) \equiv x^{(1)}\theta^{(1)} + x^{(2)}\theta^{(2)} - K, \\ &m^{(2)}(x, \theta) \equiv -\theta^{(2)}, \quad m^{(3)}(x, \theta) \equiv -\theta^{(1)}. \end{aligned}$$

We note that Assumptions 3.2, 3.4, 3.5, and 3.6 then hold provided $E[X^{(1)}] > 0$, $E[X^{(2)}] > 0$, and $B > K/\min\{E[X^{(1)}], E[X^{(2)}]\}$. Further suppose $P \ll \mu$, and $\eta \mapsto h_\eta$ is a curve in L^2_μ with

$$\begin{aligned} &\int h_\eta^2(x)\, d\mu(x) = 1, \quad \int x^{(1)} h_\eta^2(x)\, d\mu(x) = E[X^{(1)}](1 + \eta), \\ &\int x^{(2)} h_\eta^2(x)\, d\mu(x) = E[X^{(2)}], \end{aligned}$$

and $h_0 = \sqrt{dP/d\mu}$. If P_η satisfies $\sqrt{dP_\eta/d\mu} = h_\eta$, then it follows that $P_\eta \in \mathbf{P}$ for η in a neighborhood of zero. However, at the point $\bar{p} \equiv \bar{v}/\|\bar{v}\|$ with $\bar{v} \equiv (E[X^{(1)}], E[X^{(2)}])'$, we obtain that

$$\nu(\bar{p}, \Theta_0(P_\eta)) = \begin{cases} \dfrac{K}{\|\bar{v}\|}, & \text{if } \eta \geq 0, \\ \dfrac{K}{\|\bar{v}\|} \dfrac{E[X^{(1)}]}{(E[X^{(1)}] + \eta)}, & \text{if } \eta < 0, \end{cases}$$

which implies that the support function is not pathwise weak-differentiable at $\eta = 0$.[7]

[7]We are indebted to Mark Machina for this example.

REMARK 3.3: The null hypothesis that Assumption 3.6(iv) fails to hold can be recast as a null hypothesis concerning moment inequalities. Specifically, let $d \in \{0,1\}^{d_F}$, $\alpha \in \mathbf{R}^{d_F}$, and

$$\begin{aligned}\mathcal{T}_1(\theta, d, P) \equiv \sum_{i=1}^{d_F}\Bigg\{& d^{(i)}\bigg(F^{(i)}\bigg(\int m(x,\theta)\,dP(x)\bigg)\bigg)^2 \\ &+ \big(1-d^{(i)}\big)\bigg(F^{(i)}\bigg(\int m(x,\theta)\,dP(x)\bigg)\bigg)_+^2\Bigg\}, \\ \mathcal{T}_2(\theta, \alpha, d, P) \equiv \sum_{j=1}^{d_\theta}\Bigg(&\sum_{i=1}^{d_F} d^{(i)}\alpha^{(i)}\nabla F^{(i)}\bigg(\int m(x,\theta)\,dP(x)\bigg) \\ &\times \int \frac{\partial}{\partial\theta^{(j)}} m(x,\theta)\,dP(x)\Bigg)^2,\end{aligned}$$

where $(a)_+ \equiv \max\{a, 0\}$. It follows that P does not satisfy Assumption 3.6(iv) if and only if there is a $\theta \in \Theta$, $d \in \{0,1\}^{d_F}$, and $\alpha \in \mathbf{R}^{d_F}$ satisfying $\sum_i d^{(i)}(\alpha^{(i)})^2 = 1$ such that $\mathcal{T}_1(\theta, d, P) + \mathcal{T}_2(\theta, \alpha, d, P) = 0$. Though the derivation of a test of this null hypothesis is beyond the scope of this paper, we note that it is closely related to the specification testing problem examined in Bugni, Canay, and Shi (2012).

REMARK 3.4: Norm constraints such as $\|\theta\|^2 \le B$ can be accommodated by setting, for example, $F^{(i)}(\int m(x,\theta)\,dQ(x)) \equiv \|\theta\|^2 - B$ for some $1 \le i \le d_F$ and all Q. Upper or lower bound constraints on individual elements $\theta^{(i)}$ of the vector θ may be similarly imposed.

3.2.2. *Inverse Information Covariance Functional*

Before characterizing the covariance kernel of the limiting efficient process $\mathbb{G}_0$, we first introduce some additional notation. Since the moment restrictions are convex in θ, the support function

$$\nu\big(p, \Theta_0(P)\big) = \sup_{\theta\in\Theta}\langle p, \theta\rangle \quad \text{s.t.} \quad F\bigg(\int m(x,\theta)\,dP(x)\bigg) \le 0 \tag{12}$$

is the maximum of a convex program. Moreover, under our assumptions, there exist unique and finite Lagrange multipliers $\lambda(p, P)$ such that $\nu(p, \Theta_0(P))$ admits the Lagrangian representation

$$\nu\big(p, \Theta_0(P)\big) = \sup_{\theta\in\Theta}\bigg\{\langle p, \theta\rangle + \lambda(p,P)' F\bigg(\int m(x,\theta)\,dP(x)\bigg)\bigg\}. \tag{13}$$

In addition, the maximizers of (12) also solve (13), and consist of the boundary points of $\Theta_0(P)$ at which $\Theta_0(P)$ is tangent to the hyperplane $\{\theta \in \mathbf{R}^{d_\theta} : \langle p, \theta\rangle = \nu(p, \Theta_0(P))\}$. These boundary points, together with their associated Lagrange multipliers, are instrumental in characterizing the semiparametric efficiency bound.

THEOREM 3.2: *Let Assumptions 3.1–3.5 hold, define* $H(\theta) \equiv \nabla F(E[m(X_i, \theta)])$, *and, for each* $\theta_1, \theta_2 \in \Theta$, *let* $\Omega(\theta_1, \theta_2) \equiv E[(m(X_i, \theta_1) - E[m(X_i, \theta_1)])(m(X_i, \theta_2) - E[m(X_i, \theta_2)])']$. *If* $P \in \mathbf{P}$, *then*

$$\begin{aligned} &I^{-1}(p_1, p_2) \\ &\quad = \lambda(p_1, P)' H\big(\theta^*(p_1)\big)\Omega\big(\theta^*(p_1), \theta^*(p_2)\big)H\big(\theta^*(p_2)\big)'\lambda(p_2, P), \end{aligned}$$

for any $\theta^*(p_1) \in \arg\max_{\theta\in\Theta_0(P)}\langle p_1, \theta\rangle$ *and any* $\theta^*(p_2) \in \arg\max_{\theta\in\Theta_0(P)}\langle p_2, \theta\rangle$.

An important implication of Theorem 3.2 is that the semiparametric efficiency bound for estimating the support function at a particular point $\bar{p} \in \mathbb{S}^{d_\theta}$ (a scalar parameter) is

$$\operatorname{Var}\big\{\lambda(\bar{p}, P)'\nabla F\big(E\big[m\big(X_i, \theta^*(\bar{p})\big)\big]\big)m\big(X_i, \theta^*(\bar{p})\big)\big\},$$

for any $\theta^*(\bar{p}) \in \arg\max_{\theta\in\Theta_0(P)}\langle \bar{p}, \theta\rangle$. Hence, since Lagrange multipliers corresponding to nonbinding moment inequalities are zero, the semiparametric efficiency bound for $\nu(\bar{p}, \Theta_0(P))$ is the variance of a linear combination of the binding constraints at the boundary point $\theta^*(\bar{p}) \in \partial\Theta_0(P)$. Heuristically, the Lagrange multipliers represent the marginal value of relaxing the constraints in expanding the boundary of the identified set outwards in the direction $\bar{p}$, that is, in increasing the value of the support function at $\bar{p}$. Thus, the semiparametric efficiency bound is the variance of a linear combination of the binding constraints, where the weight each constraint receives is determined by its importance in shaping the boundary of the identified set at the point $\theta^*(\bar{p}) \in \partial\Theta_0(P)$.

3.3. *Related Model*

Our results are most easily extendable to settings where the identified set is also convex. To illustrate this point, we now highlight a close connection of the problem we study with an incomplete linear model previously examined in Beresteanu and Molinari (2008) and Bontemps, Magnac, and Maurin (2012).

For $Z \in \mathbf{R}^{d_Z}$, $Y \in \mathbf{R}$, and $V \in \mathbf{R}^{d_Z}$, we consider the identified set for the parameter θ_0 satisfying

$$E\big[V\big(Y - Z'\theta_0\big)\big] = 0, \tag{14}$$

when Y is not observed but is instead known to satisfy $Y_L \leq Y \leq Y_U$, with (Y_L, Y_U) observable. Letting $X \equiv (Y_L, Y_U, V', Z')'$ and P denote its distribution, we then obtain that, under appropriate moment restrictions, the identified set for θ_0 is given by

$$\Theta_{0,I}(P) \equiv \big\{\theta \in \mathbf{R}^{d_Z} : E\big[V\big(\tilde{Y} - Z'\theta\big)\big] = 0 \text{ for some r.v. } \tilde{Y} \\ \text{s.t. } Y_L \leq \tilde{Y} \leq Y_U \text{ a.s.}\big\}.$$

If $\Sigma(P) \equiv \int vz'\,dP(x)$ is invertible, then $\Theta_{0,I}(P)$ is bounded and convex with support function

$$\nu\big(p, \Theta_{0,I}(P)\big) \\ = \int p'\Sigma(P)^{-1}v\big(y_L + 1\big\{p'\Sigma(P)^{-1}v > 0\big\}(y_U - y_L)\big)\,dP(x); \tag{15}$$

see Bontemps, Magnac, and Maurin (2012). We impose that Z and V be of equal dimension because it is only in this instance that (15) holds, which greatly simplifies verifying pathwise weak-differentiability.

In order to derive an efficiency bound for estimating $\nu(\cdot, \Theta_{0,I}(P))$, we assume $P \in \mathbf{P_I}$, where

$$\mathbf{P_I} \equiv \bigg\{P \ll \mu : \int vz'\,dP(x) \text{ is invertible}\bigg\}$$

for some $\mu \in \mathbf{M}$. Unlike in Theorem 3.2, however, additional requirements are imposed on μ.

ASSUMPTION 3.7: (i) $\mathcal{X} \subset \mathbf{R}^{d_X}$ *is compact*; (ii) $\mu \in \mathbf{M}$ *satisfies* $\mu((y_L, y_U, v', z')' : y_L \leq y_U) = 1$; *and* (iii) $\mu((y_L, y_U, v', z')' : c'v = 0) = 0$ *for any vector* $c \in \mathbf{R}^{d_Z}$ *with* $c \neq 0$.

Since $P \ll \mu$ for all $P \in \mathbf{P_I}$, we note that Assumptions 3.7(i)–(ii) imply X is bounded and $Y_L \leq Y_U$ P-a.s. In particular, Y_L and Y_U must be bounded P-a.s. for all $P \in \mathbf{P_I}$, and hence all measurable selections of the random set $[Y_L, Y_U]$ are integrable. Similarly, $P \ll \mu$ and Assumption 3.7(iii) ensure $P(c'V = 0) = 0$ for all $c \neq 0$. Beresteanu and Molinari (2008) first established the importance of this requirement, showing that $\Theta_{0,I}(P)$ is strictly convex if P satisfies it, but has "flat faces" otherwise. Interestingly, in close connection to Remark 3.2, $Q \mapsto \nu(p, \Theta_{0,I}(Q))$ may not be pathwise weak-differentiable when Assumption 3.7(iii) fails to hold because the slopes of the resulting "flat faces" may then depend on P.[8]

[8]We thank Francesca Molinari for this insight; see the Supplemental Material for a more detailed discussion.

THEOREM 3.3: *Let Assumptions 3.1 and 3.7 hold, and define* $\psi_\nu : \mathbb{S}^{d_\theta} \times \mathcal{X} \to \mathbf{R}$, $\psi_\Sigma : \mathbb{S}^{d_\theta} \times \mathcal{X} \to \mathbf{R}$ *by*

$$\psi_\nu(p, x, P) \equiv \{y_L + 1\{p'\Sigma(P)^{-1}v > 0\}(y_U - y_L)\}v'\Sigma(P)^{-1}p, \tag{16}$$

$$\begin{aligned}\psi_\Sigma(p, x, P) \equiv\ & p'\Sigma(P)^{-1}zv'\Sigma(P)^{-1} \\ & \times \left\{ \int v(y_L + 1\{p'\Sigma(P)^{-1}v > 0\}(y_U - y_L))\, dP(x) \right\}.\end{aligned} \tag{17}$$

If $P \in \mathbf{P_I}$ *and* $\psi \equiv \psi_\nu - \psi_\Sigma$, *then the semiparametric efficiency bound for* $\nu(\cdot, \Theta_{0,I}(P))$ *satisfies*

$$\begin{aligned} I^{-1}(p_1, p_2) = E\big[&\big(\psi(p_1, X_i, P) - E[\psi(p_1, X_i, P)]\big) \\ &\times \big(\psi(p_2, X_i, P) - E[\psi(p_2, X_i, P)]\big)\big]. \end{aligned}$$

The semiparametric efficiency bound of Theorem 3.3 coincides with the asymptotic distribution of the estimators studied in Beresteanu and Molinari (2008) and Bontemps, Magnac, and Maurin (2012), thus verifying their efficiency. We also note that if $P(Y_L = Y_U) = 1$, so that the model is identified, then Theorem 3.3 implies that the efficient estimator is $p \mapsto \langle p, \hat{\theta}\rangle$ for $\hat{\theta}$ the GMM estimator of (14).

4. EFFICIENT ESTIMATION

4.1. *The Estimator*

Given a sample $\{X_i\}_{i=1}^n$, we let $\hat{P}_n$ denote the empirical measure; that is, $\hat{P}_n(A) \equiv \frac{1}{n}\sum_i 1\{X_i \in A\}$ for any Borel set $A \subseteq \mathcal{X}$. Under Assumption 3.1, $\hat{P}_n$ is consistent for P under the τ-topology. Therefore, a natural estimator for the support function $\nu(\cdot, \Theta_0(P))$ is its sample analog

$$\nu(p, \Theta_0(\hat{P}_n)) = \sup_{\theta\in\Theta}\langle p, \theta\rangle \quad \text{s.t.} \quad F\left(\frac{1}{n}\sum_{i=1}^n m(X_i, \theta)\right) \le 0. \tag{18}$$

It is useful to note that Assumption 3.6(ii) implies that the constraints in (18) are convex in $\theta \in \Theta$ with probability tending to 1. As a result, $\nu(p, \Theta_0(\hat{P}_n))$ also admits a characterization as a Lagrangian:

$$\nu(p, \Theta_0(\hat{P}_n)) = \sup_{\theta\in\Theta}\left\{\langle p, \theta\rangle + \lambda(p, \hat{P}_n)'F\left(\frac{1}{n}\sum_{i=1}^n m(X_i, \theta)\right)\right\}. \tag{19}$$

This dual representation, together with the envelope theorem of Milgrom and Segal (2002), enables us to conduct a Taylor expansion of $\nu(\cdot, \Theta_0(\hat{P}_n))$ around

$\nu(\cdot, \Theta_0(P))$. In this manner, we are able to characterize the influence function of $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ (in $\mathcal{C}(\mathbb{S}^{d_\theta})$), and establish its efficiency.

THEOREM 4.1: *If Assumptions 3.1, 3.2, 3.3, 3.4, and 3.5 hold and $P \in \mathbf{P}$, then it follows that*: (i) $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ *is a regular estimator for* $\nu(\cdot, \Theta_0(P))$; (ii) *uniformly in* $p \in \mathbb{S}^{d_\theta}$,

$$\begin{aligned}
&\sqrt{n}\big\{\nu\big(p, \Theta_0(\hat{P}_n)\big) - \nu\big(p, \Theta_0(P)\big)\big\} \\
&\quad = \lambda(p, P)' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H\big(\theta^*(p)\big) \\
&\qquad \times \big\{m\big(X_i, \theta^*(p)\big) - E\big[m\big(X_i, \theta^*(p)\big)\big]\big\} + o_p(1),
\end{aligned}$$

where $\theta^*(p) \in \arg\max_{\theta \in \Theta_0(P)} \langle p, \theta\rangle$ *for all* $p \in \mathbb{S}^{d_\theta}$; (iii) *as a process in* $\mathcal{C}(\mathbb{S}^{d_\theta})$,

$$\sqrt{n}\big\{\nu\big(\cdot, \Theta_0(\hat{P}_n)\big) - \nu\big(\cdot, \Theta_0(P)\big)\big\} \xrightarrow{L} \mathbb{G}_0,$$

where $\mathbb{G}_0$ *is a mean zero tight Gaussian process on* $\mathcal{C}(\mathbb{S}^{d_\theta})$ *with* $\mathrm{Cov}(\mathbb{G}_0(p_1), \mathbb{G}_0(p_2)) = I^{-1}(p_1, p_2)$.

In moment inequality models, it is common for the limiting distribution of statistics $\{T_n(\theta)\}$ to be discontinuous in $\theta \in \Theta_0(P)$. It is interesting to note that, in contrast, in Theorem 4.1 $\mathbb{G}_0$ is continuous in $p \in \mathbb{S}^{d_\theta}$ almost surely.[9] Heuristically, the continuity of $\mathbb{G}_0$ results from the Lagrange multipliers determining the weight a binding constraint receives at each $p \in \mathbb{S}^{d_\theta}$. Hence, if p_1 and p_2 are close, then the complementary slackness condition and continuity of $p \mapsto \lambda(p, P)$ imply that constraints that are binding at p_1 but not p_2 must have a correspondingly small weight. As a result, the empirical process is continuous despite different constraints being binding at different $p \in \mathbb{S}^{d_\theta}$.

4.2. *Asymptotic Risk*

Theorem 4.1 implies $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ is asymptotically optimal for a wide class of loss functions.

THEOREM 4.2: *Let Assumptions 3.1–3.5 hold*, $P \in \mathbf{P}$, *and* $L : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}_+$ *be a subconvex function*[10] *such that*, *for all* $f \in \mathcal{C}(\mathbb{S}^{d_\theta})$, $L(f) \leq M_0 + M_1 \|f\|_\infty^\kappa$ *for*

[9] A key difference being $\mathbb{G}_0$ has domain $\mathbb{S}^{d_\theta}$, while test statistics $\{T_n(\theta)\}$ often have domain Θ.

[10] L is subconvex if, for all $f \in \mathcal{C}(\mathbb{S}^{d_\theta})$, $L(0) = 0 \leq L(f)$, $L(f) = L(-f)$, and $\{f : L(f) \leq c\}$ is convex for all $c \in \mathbf{R}$.

some $M_0, M_1 > 0$ *and* $\kappa < \infty$. *If* $\mathbb{D}_0$ *are the continuity points of* L *and* $P(\mathbb{G}_0 \in \mathbb{D}_0) = 1$, *then*, *for any regular estimator* $\{T_n\}$ *of* $\nu(\cdot, \Theta_0(P))$,

$$\liminf_{n\to\infty} E\big[L\big(\sqrt{n}\big\{T_n - \nu\big(\cdot, \Theta_0(P)\big)\big\}\big)\big]$$
$$\geq \limsup_{n\to\infty} E\big[L\big(\sqrt{n}\big\{\nu\big(\cdot, \Theta_0(\hat{P}_n)\big) - \nu\big(\cdot, \Theta_0(P)\big)\big\}\big)\big] = E\big[L(\mathbb{G}_0)\big].$$

The lower bound on asymptotic risk obtained in Theorem 4.2 is a direct consequence of the Convolution Theorem and, in fact, holds for any subconvex function $L : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}_+$. The requirement that $L(f)$ be majorized by a polynomial in the norm of f is imposed to show that the plug-in estimator actually attains the bound. Below, we provide some examples of possible choices of loss function L.

EXAMPLE 4.1: Suppose in Example 2.1 we are concerned with the mean absolute error in estimating the upper bound on $E[Y|Z = z_0]$ for some $z_0 \in \mathcal{Z}$. Since $\sup_{\theta\in\Theta_0(P)}\langle z_0, \theta\rangle = \|z_0\|\nu(z_0/\|z_0\|, \Theta_0(P))$, we may apply Theorem 4.2 with $L(f) = |\|z_0\|f(z_0/\|z_0\|)|$ for any $f \in \mathcal{C}(\mathbb{S}^{d_\theta})$. Alternatively, for the expected maximal estimation error across multiple upper (or lower) bounds, we may let $L(f) = \sup_{p\in\mathbb{S}^{d_\theta}} |w(p)f(p)|$ for any bounded weight function $w : \mathbb{S}^{d_\theta} \to \mathbf{R}$.

EXAMPLE 4.2: If we are interested in the mean squared error of estimating the diameter of the identified set for a coordinate $\theta^{(i)}$ of θ, then we may set $L(f) = (f(p_0) - f(-p_0))^2$, where $p_0^{(i)} = 1$ and $p_0^{(j)} = 0$ for all $j \neq i$. Analogously, a common measure of "center" of a convex set K is given by its Steiner point, defined as $\int p\nu(p, K)\, d\Lambda(p)$ for Λ the uniform measure on $\mathbb{S}^{d_\theta}$. To obtain the mean squared error in estimating the center of $\Theta_0(P)$, we may then set $L(f) = (\int pf(p)\, d\Lambda(p))^2$.

Due to the equality of the Hausdorff distance between convex sets and the supremum distance between their corresponding support functions (see (9)), Theorem 4.2 further implies an asymptotic optimality result for asymptotic risk based on the Hausdorff metric. Specifically, define

$$\hat{\Theta}_n \equiv \mathrm{co}\big\{\Theta_0(\hat{P}_n)\big\}, \tag{20}$$

where $\mathrm{co}\{\Theta_0(\hat{P}_n)\}$ denotes the convex hull of $\Theta_0(\hat{P}_n)$. Corollary 4.1 then establishes that, for a wide class of loss functions, $\hat{\Theta}_n$ is an asymptotically optimal estimator of $\Theta_0(P)$.

COROLLARY 4.1: *Let Assumptions* 3.1–3.5 *hold*, $P \in \mathbf{P}$, *and* $L : \mathbf{R}_+ \to \mathbf{R}_+$ *be a subconvex function continuous on* $\mathbb{D}_0 \subseteq \mathbf{R}_+$, *and satisfying* $\limsup_{a\to\infty} L(a)a^{-\kappa} <$

∞ *for some* $\kappa > 0$. *If* $\{K_n\}$ *is a regular convex compact valued set estimator for* $\Theta_0(P)$, *and* $P(\|\mathbb{G}_0\|_\infty \in \mathbb{D}_0) = 1$, *then*[11]

$$\begin{aligned}\liminf_{n\to\infty} E\big[L\big(\sqrt{n}d_H\big(K_n, \Theta_0(P)\big)\big)\big] &\geq \limsup_{n\to\infty} E\big[L\big(\sqrt{n}d_H\big(\hat{\Theta}_n, \Theta_0(P)\big)\big)\big] \\ &= E\big[L\big(\|\mathbb{G}_0\|_\infty\big)\big].\end{aligned}$$

For instance, setting $L(a) = a^2$ in Corollary 4.1 yields quadratic loss based on Hausdorff distance. Alternatively, by selecting $L(a) = 1\{a \geq t\}$ for any $t \in \mathbf{R}$, we can conclude that the asymptotic distribution of $\sqrt{n}d_H(\hat{\Theta}_n, \Theta_0(P))$ is first order stochastically dominated by that of $\sqrt{n}d_H(K_n, \Theta_0(P))$.

4.3. *Marginal Identified Sets*

It is often of interest to estimate the identified set of a coordinate or subvector of θ, rather than $\Theta_0(P)$ itself. The support functions of these "marginal" identified sets are given by restrictions of $\nu(\cdot, \Theta_0(P))$ to known subsets $\mathbb{C} \subseteq \mathbb{S}^{d_\theta}$, which we denote by $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$; see Remark 4.1.[12]

In a finite dimensional setting, the coordinates of an efficient estimator are themselves efficient for the coordinates of the parameter of interest. Analogously, Theorem 4.1 implies that the restriction of the "plug-in" estimator, denoted $\{\nu_{|\mathbb{C}}(\cdot, \Theta_0(\hat{P}_n))\}$, is an efficient estimator for $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$. However, the more modest goal of obtaining an efficient estimator for $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$, rather than for $\nu(\cdot, \Theta_0(P))$, can be accomplished under less stringent assumptions on F and m. Specifically, it is possible to allow the slope of linear constraints to depend on P, provided we impose that P satisfies the following.

ASSUMPTION 4.1: *For all* $p \in \mathbb{C}$, *there is a unique* $\theta^*(p) \in \Theta_0(P)$ *with* $\langle p, \theta^*(p)\rangle = \nu(p, \Theta_0(P))$.

Assumption 4.1 imposes that, at each $p \in \mathbb{C}$, the corresponding tangent hyperplane be supported by a unique boundary point of $\Theta_0(P)$. A similar requirement is also imposed by Pakes et al. (2006) when deriving the asymptotic distribution of estimators of extremum points of the identified set. In Remark 3.2, for example, Assumption 4.1 excludes $p \in \mathbb{S}^{d_\theta}$ for which the tangent hyperplane coincides with a "flat face" of $\Theta_0(P)$—precisely the points at which $\nu(p, \Theta_0(P))$ is not pathwise weak-differentiable. To reflect this additional restriction on P, we define

$$\mathbf{P_L} \equiv \big\{P \in \mathbf{M} : P \ll \mu \text{ and Assumptions 3.6(i)–(iv) and 4.1 hold}\big\}.$$

[11]We say $\{K_n\}$ is a regular estimator of $\Theta_0(P)$ if its support function $\nu(\cdot, K_n)$ is a regular estimator for $\nu(\cdot, \Theta_0(P))$.

[12]For any subset $\mathbb{C} \subseteq \mathbb{S}^{d_\theta}$, $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P)) : \mathbb{C} \to \mathbf{R}$ is defined by $\nu_{|\mathbb{C}}(p, \Theta_0(P)) = \nu(p, \Theta_0(P))$ for all $p \in \mathbb{C}$.

To allow the slope of linear constraints to depend on P, we let $m_A : \mathcal{X} \to \mathbf{R}^{d_{m_A}}$ and

$$m(x, \theta) \equiv \big(m_S(x, \theta)', m_A(x)', \theta'\big)'. \tag{21}$$

For $v \mapsto F_A(v)$ a map such that $F_A(v)$ is a $d_F \times d_\theta$ matrix for each $v \in \mathbf{R}^{d_{m_A}}$, we then impose

$$\begin{aligned} F\Big(\int m(x, \theta)\, dP(x)\Big) \\ = F_A\Big(\int m_A(x)\, dP(x)\Big)\theta + F_S\Big(\int m_S(x, \theta)\, dP(x)\Big) \end{aligned} \tag{22}$$

(contrast to (3)). We formalize this new structure for the inequalities in the following assumption.

ASSUMPTION 4.2: (i) *The functions* $m : \mathcal{X} \times \Theta \to \mathbf{R}^{d_m}$ *and* $F : \mathbf{R}^{d_m} \to \mathbf{R}^{d_F}$ *satisfy* (21) *and* (22); (ii) *for each* $i \in \{1, \dots, d_F\}$, *we have either* $\mathcal{S}_i = \emptyset$ *or* $\mathcal{S}_i = \{1, \dots, d_\theta\}$.

Assumption 4.2(i) generalizes Assumption 3.3, since we can set $F_A(v) = A$ for all $v \in \mathbf{R}^{d_{m_A}}$ and some known $d_F \times d_\theta$ matrix A. Assumption 4.2(ii) additionally imposes that each constraint be either linear or strictly convex in θ. This requirement is not necessary for showing existence of a regular estimator of $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$, but it is needed to establish the semiparametric efficiency of $\{\nu_{|\mathbb{C}}(\cdot, \Theta_0(\hat{P}_n))\}$. Under Assumption 4.2(ii), knowledge that P satisfies Assumption 4.1 does not affect the tangent space, and hence the plug-in estimator remains efficient. In contrast, it is possible to construct examples violating Assumption 4.2(ii) where the tangent spaces relative to $\mathbf{P_L}$ and $\mathbf{P}$ differ, and hence so do the semiparametric efficiency bounds. Characterizing the efficiency bound without Assumption 4.2(ii) is a challenging problem beyond the scope of this paper.

THEOREM 4.3: *Let Assumptions* 3.1, 3.2, 3.4, 3.5, *and* 4.2 *hold*. *If* $P \in \mathbf{P_L}$ *and* $\mathbb{C} \subseteq \mathbb{S}^{d_\theta}$ *is compact*, *then* $\{\nu_{|\mathbb{C}}(\cdot, \Theta_0(\hat{P}_n))\}$ *is a semiparametrically efficient estimator of* $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$ (*in* $\mathcal{C}(\mathbb{C})$).

REMARK 4.1: Suppose $\theta = (\theta_1, \theta_2) \in \mathbf{R}^{d_{\theta_1}+d_{\theta_2}}$, and we are interested in the marginal identified set

$$\Theta_{0,M}(P) \equiv \big\{\theta_1 \in \mathbf{R}^{d_{\theta_1}} : (\theta_1, \theta_2) \in \Theta_0(P) \text{ for some } \theta_2 \in \mathbf{R}^{d_{\theta_2}}\big\}.$$

For any $p_1 \in \mathbb{S}^{d_{\theta_1}}$, the support function of the marginal identified set $\Theta_{0,M}(P)$ then satisfies

$$\nu\big(p_1, \Theta_{0,M}(P)\big) = \sup_{\theta_1 \in \Theta_{0,M}(P)} \langle p_1, \theta_1 \rangle = \sup_{(\theta_1,\theta_2)\in\Theta_0(P)} \big\{\langle p_1, \theta_1\rangle + \langle 0, \theta_2\rangle\big\}$$
$$= \nu\big((p_1, 0), \Theta_0(P)\big).$$

Hence, we obtain $\nu(\cdot, \Theta_{0,M}(P)) = \nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$ for $\mathbb{C} \equiv \{(p_1, p_2) \in \mathbb{S}^{d_{\theta_1}+d_{\theta_2}} : p_2 = 0\}$.

5. A CONSISTENT BOOTSTRAP

We obtain a consistent bootstrap procedure by following a "score based" approach as proposed in Lewbel (1995); see also Donald and Hsu (2009) and Kline and Santos (2012). In particular, for $W_i \in \mathbf{R}$ a mean zero random variable and $\{W_i\}_{i=1}^n$ an i.i.d. sample independent of $\{X_i\}_{i=1}^n$, we let

$$G_n^*(p) \equiv \lambda(p, \hat{P}_n)' \nabla F\left(\frac{1}{n}\sum_{i=1}^n m\big(X_i, \hat{\theta}(p)\big)\right) \times \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{m\big(X_i, \hat{\theta}(p)\big) - \frac{1}{n}\sum_{i=1}^n m\big(X_i, \hat{\theta}(p)\big)\right\} W_i, \tag{23}$$

where $\lambda(p, \hat{P}_n)$ is as in (19) and $\hat{\theta}(p)$ is any maximizer for the optimization problem in (19). Heuristically, the stochastic process $p \mapsto G_n^*(p)$ is constructed by perturbing an estimate of the efficient influence function (or score) by the random weights $\{W_i\}_{i=1}^n$. These weights are assumed to satisfy the following.

ASSUMPTION 5.1: (i) $\{X_i, W_i\}_{i=1}^n$ *is an i.i.d. sample*; (ii) W_i *is independent of* X_i; (iii) W_i *satisfies* $E[W_i] = 0$, $E[W_i^2] = 1$ *and* $E[|W_i|^{2+\delta}] < \infty$ *for some* $\delta > 0$.

By construction, the distribution of G_n^* depends on that of both $\{X_i\}_{i=1}^n$ and $\{W_i\}_{i=1}^n$. We show, however, that the distribution of G_n^* *conditional* on the data $\{X_i\}_{i=1}^n$ (but not $\{W_i\}_{i=1}^n$) is a consistent estimator for the law of $\mathbb{G}_0$. Formally, letting L^* denote a law statement conditional on $\{X_i\}_{i=1}^n$, Theorem 5.1 establishes consistency of the law of G_n^* under L^* for that of $\mathbb{G}_0$.

THEOREM 5.1: *If Assumptions* 3.1–3.5 *and* 5.1 *hold and* $P \in \mathbf{P}$, *then* $G_n^* \xrightarrow{L^*} \mathbb{G}_0$ (*in probability*).

5.1. *Estimating Critical Values*

In order to conduct inference, it is often necessary to estimate quantiles of transformations of $\mathbb{G}_0$. In this section, we develop a procedure applicable when the transformation is of the form

$$\sup_{p\in\Psi_0} \Upsilon\big(\mathbb{G}_0(p)\big), \tag{24}$$

where $\Psi_0 \subseteq \mathbb{S}^{d_\theta}$ and $\Upsilon : \mathbf{R} \to \mathbf{R}$ is a known continuous function. The set $\Psi_0 \subseteq \mathbb{S}^{d_\theta}$ need not be known, but we assume the availability of a consistent estimator $\{\hat{\Psi}_n\}$ for Ψ_0 in Hausdorff distance.

ASSUMPTION 5.2: (i) $\Upsilon : \mathbf{R} \to \mathbf{R}$ *is continuous*; (ii) $\hat{\Psi}_n \subseteq \mathbb{S}^{d_\theta}$ *is compact almost surely*; (iii) $\{\hat{\Psi}_n\}$ *satisfies* $d_H(\hat{\Psi}_n, \Psi_0) = o_p(1)$ *with* Ψ_0 *compact*.

Quantiles of random variables as in (24) may then be consistently estimated employing

$$\hat{c}_{1-\alpha} \equiv \inf\Big\{c : P\Big(\sup_{p\in\hat{\Psi}_n} \Upsilon\big(G_n^*(p)\big) \le c \Big| \{X_i\}_{i=1}^n\Big) \ge 1-\alpha\Big\}. \tag{25}$$

While $\hat{c}_{1-\alpha}$ is often not explicitly computable, it can be approximated using the following algorithm:

STEP 1: Compute the full sample support function estimate $\nu(\cdot, \Theta_0(\hat{P}_n))$ and obtain the Lagrange multipliers $\{\lambda(p, \hat{P}_n)\}_{p\in\mathbb{S}^{d_\theta}}$ and corresponding maximizers $\{\hat{\theta}(p)\}_{p\in\mathbb{S}^{d_\theta}}$ to (19).

STEP 2: Generate B samples $\{W_i\}_{i=1}^n$ to construct a sample of size B of G_n^* (denoted $\{G_{n,b}^*\}_{b=1}^B$).

STEP 3: Approximate $\hat{c}_{1-\alpha}$ by the $1-\alpha$ quantile of $\{\sup_{p\in\hat{\Psi}_n} \Upsilon(G_{n,b}^*(p))\}_{b=1}^B$.

As Theorem 5.2 establishes, $\hat{c}_{1-\alpha}$ is indeed consistent for the desired quantile.

THEOREM 5.2: *Let Assumptions 3.1–3.5, 5.1, 5.2 hold and* $P \in \mathbf{P}$. *If the cumulative distribution function* (*cdf*) *of* $\sup_{p\in\Psi_0} \Upsilon(\mathbb{G}_0(p))$ *is continuous and strictly increasing at its* $1-\alpha$ *quantile*, *denoted* $c_{1-\alpha}$, *then* $\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}$.

Theorem 5.2 may be employed, for example, to construct confidence regions for $\Theta_0(P)$.

EXAMPLE 5.1: Let $\hat{\Theta}_n^\varepsilon \equiv \{\theta \in \mathbf{R}^{d_\theta} : \inf_{\tilde{\theta}\in\hat{\Theta}_n} \|\theta - \tilde{\theta}\| \leq \varepsilon\}$, and $c_{1-\alpha}$ denote the $1-\alpha$ quantile of $\sup_{p\in\mathbb{S}^{d_\theta}}(-\mathbb{G}_0(p))_+$. Beresteanu and Molinari (2008) then established that

$$\lim_{n\to\infty} P\big(\Theta_0(P) \subseteq \hat{\Theta}_n^{\hat{c}_{1-\alpha}/\sqrt{n}}\big) = 1-\alpha \tag{26}$$

for any consistent estimator $\hat{c}_{1-\alpha}$ for $c_{1-\alpha}$. In particular, by letting $\Upsilon(a) = (-a)_+$, and $\hat{\Psi}_n = \Psi_0 = \mathbb{S}^{d_\theta}$, Theorem 5.2 implies that (26) holds if $c_{1-\alpha}$ is estimated employing the proposed bootstrap. Alternatively, Chernozhukov, Kocatulum, and Menzel (2012) provided a related construction based on the efficient estimator that is equivariant to transformations of the parameters.

5.2. *Application to Testing*

As an illustration of the applicability of Theorem 5.2, we consider the hypothesis testing problem

$$H_0 : \theta \in \Theta_0(P), \quad H_1 : \theta \notin \Theta_0(P), \tag{27}$$

which is commonly inverted to construct confidence regions that cover each element of $\Theta_0(P)$ with a prespecified probability. In a related setting, Kaido (2012) tested (27) employing the statistic[13]

$$J_n(\theta) \equiv \sqrt{n}\vec{d}_H\big(\{\theta\}, \hat{\Theta}_n\big). \tag{28}$$

For $\mathfrak{M}(\theta) \equiv \arg\max_{p\in\mathbb{S}^{d_\theta}}\{\nu(p,\{\theta\}) - \nu(p,\Theta_0(P))\}$, the appropriate critical value for $J_n(\theta)$ is then

$$c_{1-\alpha}(\theta) \equiv \inf\Big\{c : P\Big(\sup_{p\in\mathfrak{M}(\theta)}\big(-\mathbb{G}_0(p)\big)_+ \leq c\Big) \geq 1-\alpha\Big\}. \tag{29}$$

Estimating $c_{1-\alpha}(\theta)$ requires a consistent estimator for $\mathfrak{M}(\theta)$, for which Kaido (2012) proposed

$$\hat{\mathfrak{M}}_n(\theta) \equiv \Big\{p \in \mathbb{S}^{d_\theta} : \big\{\nu\big(p,\{\theta\}\big) - \nu\big(p,\Theta_0(\hat{P}_n)\big)\big\} \geq \sup_{\tilde{p}\in\mathbb{S}^{d_\theta}}\big\{\nu\big(\tilde{p},\{\theta\}\big) - \nu\big(\tilde{p},\Theta_0(\hat{P}_n)\big)\big\} - \frac{\kappa_n}{\sqrt{n}}\Big\}, \tag{30}$$

[13]Kaido (2012) examined an arbitrary estimator of $\nu(\cdot,\Theta_0(P))$, not necessarily the efficient one. This type of test statistic was first studied by Bontemps, Magnac, and Maurin (2012) in the context of the incomplete linear model of Section 3.3.

which satisfies $d_H(\mathfrak{M}(\theta), \hat{\mathfrak{M}}_n(\theta)) = o_p(1)$ provided $\kappa_n = o(n^{1/2})$ and $\kappa_n \uparrow \infty$. Applying Theorem 5.2 with $\Upsilon(a) = (-a)_+$, $\Psi_0 = \mathfrak{M}(\theta)$, and $\hat{\Psi}_n = \hat{\mathfrak{M}}_n(\theta)$ then implies that a consistent estimate of $c_{1-\alpha}(\theta)$ is

$$\hat{c}_{1-\alpha}(\theta) \equiv \inf\Big\{c : P\Big(\sup_{p \in \hat{\mathfrak{M}}_n(\theta)} \big(-G_n^*(p)\big)_+ \leq c \big| \{X_i\}_{i=1}^n\Big) \geq 1 - \alpha\Big\}. \tag{31}$$

Theorem 5.3 establishes that the proposed bootstrap delivers pointwise (in P) asymptotic size control.

THEOREM 5.3: *Let Assumptions* 3.1–3.5 *and* 5.1 *hold*, $P \in \mathbf{P}$, $\alpha \in (0, 0.5)$, *and* $\kappa_n \uparrow \infty$ *with* $\kappa_n = o(n^{1/2})$. *If* $\theta \in \Theta_0(P)$, *and* $\mathrm{Var}\{\mathbb{G}_0(p)\} > 0$ *for some* $p \in \mathfrak{M}(\theta)$, *then it follows that*

$$\liminf_{n \to \infty} P\big(J_n(\theta) \leq \hat{c}_{1-\alpha}(\theta)\big) \geq 1 - \alpha. \tag{32}$$

5.2.1. *Local Properties*

The test that rejects (27) whenever $J_n(\theta) > \hat{c}_{1-\alpha}(\theta)$ satisfies a local optimality property. Specifically, we show that the power function of any test that controls size over local parametric submodels must be weakly smaller than that of a test based on $J_n(\theta)$ for all $\theta \in \partial\Theta_0(P)$ that are supported by a unique hyperplane. Formally, let $h_\eta = \sqrt{dP_\eta/d\mu}$ and $\mathbf{H}(\theta)$ denote the set of submodels $\eta \mapsto P_\eta$ in $\mathbf{P}$ with

$$\begin{aligned} &\text{(i)} && h_0 = \sqrt{dP/d\mu}, \\ &\text{(ii)} && \theta \in \Theta_0(P_\eta) \quad \text{if} \quad \eta \leq 0, \\ &\text{(iii)} && \theta \notin \Theta_0(P_\eta) \quad \text{if} \quad \eta > 0. \end{aligned} \tag{33}$$

Thus, $\mathbf{H}(\theta)$ is the set of submodels passing through P for which P_η satisfies the null hypothesis in (27) for $\eta \leq 0$, and the alternative for $\eta > 0$. We consider tests in terms of their power functions $\pi : \mathbf{H}(\theta) \to [0, 1]$, where $\pi(P_\eta)$ is the probability that the null hypothesis is rejected when $X_i \sim P_\eta$.

THEOREM 5.4: *Let Assumptions* 3.1–3.5 *and* 5.1 *hold*, $P \in \mathbf{P}$, $\theta_0 \in \partial\Theta_0(P)$ *with* $\mathfrak{M}(\theta_0) = \{p_0\}$ *and* $\mathrm{Var}\{\mathbb{G}_0(p_0)\} > 0$, *and* $\{\pi_n\}$ *be any sequence of power functions such that*, *for any* $P_\eta \in \mathbf{H}(\theta_0)$, $\eta \leq 0$,

$$\limsup_{n \to \infty} \pi_n(P_{\eta/\sqrt{n}}) \leq \alpha. \tag{34}$$

If $\{\pi_n^*\}$ *is the power function of the test that rejects when* $J_n(\theta_0) > \hat{c}_{1-\alpha}(\theta_0)$, *then* $\{\pi_n^*\}$ *satisfies* (34). *Moreover*, *for* $\tilde{l}(x) \equiv -\lambda(p_0, P)' H(\theta_0)\{m(x, \theta_0) -$

$E[m(X_i, \theta_0)]\}$ and any $P_\eta \in \mathbf{H}(\theta_0)$, $\eta > 0$,

$$\begin{aligned}(35)\qquad \limsup_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) &\le \lim_{n\to\infty} \pi_n^*(P_{\eta/\sqrt{n}}) \\ &= 1 - \Phi\left(z_{1-\alpha} - \eta\frac{2E[\tilde{l}(X_i)\dot{h}_0(X_i)/h_0(X_i)]}{\sqrt{E[\mathbb{G}_0^2(p_0)]}}\right),\end{aligned}$$

where Φ is the cdf of a standard normal random variable and $z_{1-\alpha}$ is its $1-\alpha$ quantile.

The null hypothesis in (27) holds if and only if $\langle p, \theta\rangle \le \nu(p, \Theta_0(P))$ for all $p \in \mathbb{S}^{d_\theta}$. When $\mathfrak{M}(\theta) = \{p_0\}$, such inequality holds with equality only at p_0. Heuristically, any local perturbation $P_{\eta/\sqrt{n}}$ of P that violates the null hypothesis in (27) must then satisfy $\langle p_0, \theta\rangle > \nu(p_0, \Theta_0(P_{\eta/\sqrt{n}}))$. As a result, it is possible to locally relate (27) to the problem of testing $\langle p_0, \theta\rangle \le \nu(p_0, \Theta_0(P))$ against $\langle p_0, \theta\rangle > \nu(p_0, \Theta_0(P))$. The limiting experiment of the latter hypothesis is akin to a one sided test for a mean, and Theorem 5.4 follows by showing that the proposed test is optimal in this context. We note, however, that the size control requirement in (34) is local to a $P \in \mathbf{P}$, and the proposed test does not necessarily control size uniformly over a larger set of distributions.

6. CONCLUSION

This paper obtains conditions under which the support function of the identified set is a regular parameter, and characterizes the semiparametric efficiency bound for estimating it. These conditions are instructive in also determining the sources of irregularity. As in standard maximum likelihood, however, the results are local in nature. Consequently, care should be taken in implementation whenever there is reason to doubt the relevance of the assumption $P \in \mathbf{P}$.

REFERENCES

AI, C., AND X. CHEN (2009): "Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions Containing Unknown Functions," Working Paper, Yale University. [389]

ANDREWS, D. W. K., AND P. J. BARWICK (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80, 2805–2826. [389]

ANDREWS, D. W. K., AND P. GUGGENBERGER (2009): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709. [389]

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157. [388]

BERESTEANU, A., AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814. [387-389,400-402,409]

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp Identification Regions in Models With Convex Predictions: Games, Individual Choice, and Incomplete Data," *Econometrica*, 79, 1785–1821. [387]

BICKEL, P. J., C. A. KLASSEN, Y. RITOV, AND J. A. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer. [394,395]

BLUNDELL, R., AND T. MACURDY (1999): "Labor Supply: A Review of Alternative Approaches," in *Handbook of Labor Economics*, Vol. 3A, ed. by O. Ashenfelter and D. Card. Amsterdam: North-Holland. [393]

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2012): "Set Identified Linear Models," *Econometrica*, 80, 1129–1155. [387,388,400-402,409]

BROWN, B. W., AND W. K. NEWEY (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66, 453–464. [389]

BUGNI, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753. [389]

BUGNI, F. A., I. A. CANAY, AND X. SHI (2012): "Specification Test in Partially Identified Models Defined by Moment Inequalities," Working Paper, Northwestern University. [399]

CANAY, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156, 408–425. [389]

CHAMBERLAIN, G. (1987): "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334. [389]

——— (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567–596. [389]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [388]

CHERNOZHUKOV, V., E. KOCATULUM, AND K. MENZEL (2012): "Inference on Sets in Finance," Working Paper, Massachusetts Institute of Technology. [409]

CHESHER, A. (2009): "Single Equation Endogenous Binary Response Models," Working Paper, Centre for Microdata Methods and Practice (CEMMAP). [388]

CILIBERTO, F., AND E. TAMER (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828. [387]

DONALD, S. G., AND Y. HSU (2009): "Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models," Working Paper, University of Texas–Austin. [407]

HANSEN, L., J. HEATON, AND E. LUTTMER (1995): "Econometric Evaluation of Asset Pricing Models," *Review of Financial Studies*, 8, 237–274. [392]

HIRANO, K., AND J. R. PORTER (2012): "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80, 1769–1790. [389,390,398]

KAIDO, H. (2012): "A Dual Approach to Inference for Partially Identified Econometric Models," Working Paper, Boston University. [388,409]

KAIDO, H., AND A. SANTOS (2014): "Supplement to 'Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities'," *Econometrica Supplemental Material*, 82, http://www.econometricsociety.org/ecta/supmat/10017_proofs.pdf; http://www.econometricsociety.org/ecta/supmat/10017_data_and_programs.zip. [389]

KLINE, P., AND A. SANTOS (2012): "A Score Based Approach to Wild Bootstrap Inference," *Journal of Econometric Methods*, 1, 23–41. [407]

LEWBEL, A. (1995): "Consistent Nonparametric Hypothesis Tests With an Application to Slutsky Symmetry," *Journal of Econometrics*, 67, 379–401. [407]

LUTTMER, E. (1996): "Asset Pricing in Economies With Frictions," *Econometrica*, 64, 1439–1467. [392]

MAHAJAN, A. (2003): "Misclassified Regressors in Binary Choice Models," Working Paper, UCLA. [388]

MANSKI, C. F. (2003): *Partial Identification of Probability Distributions*. New York: Springer. [387]

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions With Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546. [387,391]

MENZEL, K. (2009): "Consistent Estimation With Many Moment Inequalities," Working Paper, New York University. [389]

MILGROM, P., AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601. [402]

PAKES, A. (2010): "Alternative Models for Moment Inequalities," *Econometrica*, 78, 1783–1822. [387,392]

PAKES, A., J. PORTER, K. HO, AND J. ISHII (2006): "Moment Inequalities and Their Application," Working Paper, Harvard University. [387,392,405]

ROMANO, J. P., AND A. M. SHAIKH (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference—Special Issue in Honor of Ted Anderson*, 138, 2786–2807. [389]

——— (2010): "Inference for the Identified set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211. [388]

ROSEN, A. M. (2008): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117. [389]

SONG, K. (2010): "Robust Estimation of Some Nonregular Parameters," Working Paper 10-20, PIER. [389]

——— (2012): "Point Decisions for Interal-Identified Parameters," Working Paper, UBC. [388]

VAN DER VAART, A. W. (1991): "On Differentiable Functionals," *The Annals of Statistics*, 19, 178–204. [390]

Dept. of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, U.S.A.; hkaido@bu.edu

and

Dept. of Economics, University of California—San Diego, 9500 Gilman Drive MS 0508, La Jolla, CA 92093-0508, U.S.A.; a2santos@ucsd.edu.

Manuscript received May, 2011; final revision received June, 2013.