

Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities

Hiroaki Kaido

Department of Economics

Boston University

hkaido@bu.edu

Andres Santos*

Department of Economics

UC San Diego

a2santos@ucsd.edu

April, 2011

Abstract

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 for \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 class of asymptotic loss functions based on Hausdorff distance. We conclude by obtaining an efficient estimator and devising a consistent bootstrap procedure for its limiting distribution. Employing these results we are able to use the efficient estimator to construct confidence regions for the identified set. A Monte Carlo study examines finite sample performance.

KEYWORDS: Semiparametric efficiency, partial identification, moment inequalities.

*We would like to thank Mark Machina and seminar participants at Boston College, Boston University, Northwestern, New York University, Wisconsin and Yale for comments that helped greatly improve this paper.

1 Introduction

In a large number of estimation problems, the data available to the researcher fails to point identify the parameter of interest but is 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 et al., 2009), certain forms of censoring (Manski and Tamer, 2002) and optimal behavior by agents in discrete choice problems (Pakes et al., 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 variable with distribution P_0 , $\Theta \subset \mathbf{R}^{d_\theta}$ denote the parameter space and $m : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^{d_m}$ and $F : \mathbf{R}^{d_m} \rightarrow \mathbf{R}^{d_F}$ be known functions. In many models, the identified set is then of the general form:

$$\Theta_0(P_0) \equiv \left\{ \theta \in \Theta : F\left(\int m(x, \theta) dP_0(x)\right) \leq 0 \right\} . \quad (1)$$

A prevalent specification is one in which F is the identity mapping, in which case this framework reduces to the traditional moment inequalities model studied in Chernozhukov et al. (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), as well as conditional moment inequalities for discrete conditioning variables.

In this paper, we contribute to the existing literature by proposing and 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_0 . 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 (Rockafellar, 1970). As a result of this embedding, estimation of a convex set may be accomplished through the estimation of its support function – a relationship that has been previously exploited by Bontemps et al. (2007), Beresteanu and Molinari (2008) and Kaido (2010).

We provide conditions under which, despite being an infinite dimensional parameter, the support function of the identified set admits for \sqrt{n} -consistent regular estimators.¹ 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 to 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 – i.e.

¹The support function of a set in \mathbf{R}^{d_θ} ($d_\theta > 1$) is an infinite dimensional object. In Section 3.1 we briefly review support functions, regularity, and semiparametric efficient estimation of infinite dimensional parameters.

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 for the support function of the identified set by deriving the covariance kernel of the efficient Gaussian process \mathbb{G}_0 .

Among the implications of semiparametric efficiency, is that an efficient estimator minimizes diverse measures of asymptotic risk among regular estimators. Moreover, 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 the Hausdorff distance to the identified set.

Having characterized the semiparametric efficiency bound, we establish that the support function of the sample analogue to (1) is in fact the efficient estimator. Despite its simplicity and efficiency, to the best of our knowledge this “plug-in” estimator for the support function of the identified set has not been previously used for estimation or inference in the context of moment inequalities. Building on an approach developed in Kline and Santos (2010), we additionally construct a simple bootstrap procedure for consistently estimating the distribution of the limiting process \mathbb{G}_0 . These results enable us to employ the efficient estimator to construct one or two sided confidence regions for the identified set, as well as confidence regions for the parameter (Imbens and Manski, 2004). Both the estimator and bootstrap procedure can be implemented at low computation cost, which we illustrate in a simulation study based on a regression model with interval censored outcome (Manski and Tamer, 2002).

In related work, Beresteanu and Molinari (2008) first employ the relationship between convex sets and their support functions for studying partially identified models. The authors derive methods for conducting inference on convex sets through their support functions, providing insights we rely upon in our analysis. The use of support functions to characterize semiparametric efficiency is, however, novel to this paper. Other work on estimation includes Hirano and Porter (2009), who establish no regular estimators exist in intersection bounds models for scalar valued parameters and Song (2010), who proposes robust estimators for such problems. Our results complement theirs by shedding light on what the sources of irregularity are in the common setting where the parameter of interest has dimension greater than one.

A large literature on the traditional moment inequalities model has focused on inference rather than estimation. The framework we employ is not as general as the one pursued in these papers which, for example, do not impose convexity; see Andrews and Jia (2008), Romano and Shaikh (2008), Andrews and Guggenberger (2009), Rosen (2009), Menzel (2009), Bugni (2010) and Canay (2010) among others. On the other hand, the additional structure not only enables us to study asymptotic efficiency, but also to employ a simpler bootstrap procedure for estimating critical

values. This paper is also part of the larger literature on efficient estimation in econometrics, which has largely focused on finite dimensional parameters identified by moment equality restrictions; see Chamberlain (1987, 1992), Brown and Newey (1998), Ai and Chen (2009), Chen and Pouzo (2009) and references therein.

The remainder of the paper is organized as follows. Section 2 introduces the moment inequalities we study and provides examples of models that fall within its scope. In Section 3 we review efficiency analysis for infinite dimensional parameters and characterize the semiparametric efficiency bound. Section 4 establishes the efficiency of the plug-in estimator and derives a consistent bootstrap procedure. In Section 5 we employ these results to obtain confidence regions, while in Section 6 we examine their performance in a Monte Carlo study. All proofs are relegated to the Appendix.

2 General Setup

We begin by introducing the restrictions on the functions F and m that will enable us to derive a semiparametric efficiency bound for the support function of the identified set. Throughout, for any vector w we let $w^{(i)}$ denote its i^{th} coordinate.

2.1 The Model

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:

$$\Theta_0(Q) \equiv \left\{ \theta \in \Theta : F\left(\int m(x, \theta) dQ(x)\right) \leq 0 \right\}. \quad (2)$$

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 $\Theta_0(P_0)$ as the value the known mapping $Q \mapsto \Theta_0(Q)$ takes at the unknown distribution P_0 .

Our analysis focuses on settings where the identified set is convex, which we ensure by requiring that the functions $\theta \mapsto F^{(i)}(\int m(x, \theta) dP_0(x))$ be themselves convex for all $1 \leq i \leq d_F$. Unfortunately, convexity is not sufficient for establishing that $\Theta_0(P_0)$ admits for a regular estimator.² 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_0 may lead to a non-differentiable change in the identified set. This lack of differentiability in turn implies there exist no asymptotically linear regular estimators (van der Vaart, 1991; Hirano and Porter, 2009).

For this reason, we assume the vector of constraints consists of two groups: constraints that are

²In Section 3.1, we discuss regular estimators and the appropriate differentiability requirement on an infinite dimensional parameter that enables us to obtain a semiparametric efficiency bound.

strictly convex in θ and constraints that are linear in θ . Specifically, define $m : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^{d_m}$ by:

$$m(x, \theta) \equiv (m_S(x, \theta)', m_L(x)', \theta' A')', \quad (3)$$

where $m_S : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^{d_{m_S}}$ and $m_L : \mathcal{X} \rightarrow \mathbf{R}^{d_{m_L}}$ are known measurable functions and A is a $d_{F_L} \times d_\theta$ matrix. For functions $F_S : \mathbf{R}^{d_{m_S}} \rightarrow \mathbf{R}^{d_{F_S}}$ and $F_L : \mathbf{R}^{d_{m_L}} \rightarrow \mathbf{R}^{d_{F_L}}$, we then let:

$$F\left(\int m(x, \theta) dP_0(x)\right) = \begin{pmatrix} F_S\left(\int m_S(x, \theta) dP_0(x)\right) \\ A\theta - F_L\left(\int m_L(x) dP_0(x)\right) \end{pmatrix}, \quad (4)$$

where $\theta \mapsto F_S^{(i)}\left(\int m_S(x, \theta) dP_0(x)\right)$ is required to be strictly convex in θ for all $1 \leq i \leq d_{F_S}$. Thus, we allow for general forms of strictly convex constraints, but we require the slope of the linear constraints to be known in order to ensure the existence of regular estimators. We will discuss the importance of this restriction in greater detail in Remark 3.1.

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 $\Theta_0(Q)$. Towards this end, for each $\theta \in \Theta_0(Q)$, we define:

$$\mathcal{A}(\theta, Q) \equiv \{i \in \{1, \dots, d_F\} : F^{(i)}\left(\int m(x, \theta) dQ(x)\right) = 0\}. \quad (5)$$

That is, for each $\theta \in \Theta_0(Q)$, $\mathcal{A}(\theta, Q)$ denotes the set of indices i for which the inequality constraints hold with equality. If θ is in the interior of $\Theta_0(Q)$, all inequalities hold strictly and hence $\mathcal{A}(\theta, Q) = \emptyset$.

2.2 Examples

In order to fix ideas, we briefly discuss applications of the general framework introduced in Section 2.1. For ease of exposition, we base our examples on simplifications of well known models.

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:

$$Y = Z'\theta_0 + \epsilon, \quad (6)$$

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

$$E[Y_L|Z = z_i] - z_i'\theta \leq 0, \quad i = 1, \dots, K \quad (7)$$

$$z_i'\theta - E[Y_U|Z = z_i] \leq 0, \quad i = 1, \dots, K. \quad (8)$$

Let $x \equiv (y_L, y_U, z)'$, $1_{\mathcal{Z}}(z) \equiv (1\{z = z_1\}, \dots, 1\{z = z_K\})'$ and $m_L(x) = (y_L 1_{\mathcal{Z}}(z)', y_U 1_{\mathcal{Z}}(z)', 1_{\mathcal{Z}}(z)')'$. Setting $A \equiv (-z_1, \dots, -z_K, z_1, \dots, z_K)'$ and $F_L : \mathbf{R}^{3K} \rightarrow \mathbf{R}^{2K}$ to be pointwise given by:

$$F_L^{(i)}(v) = \begin{cases} -\frac{v^{(i)}}{v^{(2K+i)}}, & i = 1, \dots, K \\ \frac{v^{(i)}}{v^{(K+i)}}, & i = K + 1, \dots, 2K, \end{cases} \quad (9)$$

it then follows that the moment inequalities in (7)-(8) can be written in the form of (4) with no strictly convex constraints. We also note that a similar construction may be applied to obtain bounds in the IV model for binary outcome studied in Chesher (2009). ■

A prominent application of moment inequality models is in the context of discrete choice; see for example Pakes et al. (2006), Pakes (2010) and references therein. The following example explores the common specification in which the inequalities are linear in the parameter of interest.

Example 2.2 (Discrete choice). Suppose an agent chooses $Z \in \mathbf{R}^{d_Z}$ from a set $\mathcal{Z} \equiv \{z_1, \dots, z_K\}$ in order 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:

$$E[\pi(Y, z, \theta_0) - \pi(Y, z^*, \theta_0)|\mathcal{F}] \leq 0 \quad (10)$$

for all $z \in \mathcal{Z}$. A common specification is that $\pi(y, z, \theta_0) = \psi(y, z) + z'\theta_0 + \nu$ for some unobservable error ν ; see Pakes et al. (2006) and Pakes (2010). Therefore, under suitable assumptions on the agent's beliefs,³ the optimality conditions in (10) then imply θ_0 must satisfy the moment inequalities:

$$E[(\psi(Y, z_j) - \psi(Y, z_i)) - (z_j - z_i)'\theta]1\{Z^* = z_i\} \leq 0 \quad (11)$$

for any $z_i, z_j \in \mathcal{Z}$. Proceeding as in Example 2.1 it is then straightforward to rewrite the restrictions in (11) in the form of (4) with no strictly convex constraints. ■

Convex moment inequalities also arise in asset pricing (Hansen et al., 1995). The following example is based on Luttmer's (1996) analysis of asset pricing models with market frictions.⁴

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 $P \in \mathbf{R}^{d_Z}$. 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:

$$E\left[\frac{1}{1+\rho}Y^{-\gamma}Z - P\right] \leq 0, \quad (12)$$

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. Letting $x \equiv (y, p, z)'$, $\theta \equiv (\rho, \gamma)'$ and defining:

$$m_S(x, \theta) \equiv \frac{1}{1+\rho}y^{-\gamma}z - p, \quad (13)$$

it follows that $\theta \mapsto E[m_S^{(i)}(X, \theta)]$ is strictly convex provided $Z^{(i)} \geq 0$ almost surely and $Z^{(i)} > 0$ with positive probability. Therefore, we may express the moment inequalities in (12) in the form of (4) with $F_s(v) = v$ and no linear constraints. ■

³See Pakes (2010) for discussion.

⁴We note our semiparametric efficiency bound is for iid 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.

3 Semiparametric Efficiency

We derive in this section the semiparametric efficiency bound for estimating the support function of the identified set $\Theta_0(P_0)$. Before proceeding, we briefly review properties of support functions as well as the theory of semiparametric efficiency bounds for infinite dimensional parameters.

3.1 Preliminaries

3.1.1 Support Functions

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 $\|p\| = \langle p, p \rangle^{\frac{1}{2}}$ be the Euclidean norm. Following the literature, we employ the Hausdorff distance to evaluate distance between sets in \mathbf{R}^{d_θ} . Hence, for any closed sets A and B we let:

$$d_H(A, B) \equiv \max\{\vec{d}_H(A, B), \vec{d}_H(B, A)\} \quad \vec{d}_H(A, B) \equiv \sup_{a \in A} \inf_{b \in B} \|a - b\|, \quad (14)$$

where d_H and \vec{d}_H are the Hausdorff and directed Hausdorff distances respectively. Note that $\vec{d}_H(A, B)$ is the distance between B and the point in A that is furthest away from it. Therefore, $A \subseteq B$ if and only if $\vec{d}_H(A, B)$ equals zero.

It is often fruitful to study sets as functions instead. When the set under consideration is compact and convex, a natural approach is to identify it with its support function. Formally, for $\mathbb{S}^{d_\theta} = \{p \in \mathbf{R}^{d_\theta} : \|p\| = 1\}$ the unit sphere in \mathbf{R}^{d_θ} , we denote by $\mathcal{C}(\mathbb{S}^{d_\theta})$ the space of all bounded continuous functions on \mathbb{S}^{d_θ} and equip $\mathcal{C}(\mathbb{S}^{d_\theta})$ with the usual supremum norm $\|f\|_\infty \equiv \sup_{p \in \mathbb{S}^{d_\theta}} |f(p)|$. For any compact convex set $K \subset \mathbf{R}^{d_\theta}$, the support function of K is then pointwise defined by:

$$\nu(p, K) \equiv \sup_{k \in K} \langle p, k \rangle. \quad (15)$$

Heuristically, the support function assigns to each vector p the signed distance between the origin and the orthogonal hyperplane that is tangent to K .⁵ 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:

$$d_H(K_1, K_2) = \sup_{p \in \mathbb{S}^{d_\theta}} |\nu(p, K_1) - \nu(p, K_2)|, \quad (16)$$

see Li et al. (2002). Therefore, convex compact sets can be identified in a precise sense with a unique element of $\mathcal{C}(\mathbb{S}^{d_\theta})$ in a way that preserves distances – i.e. there exists an isometric embedding.

In our analysis, we are interested in the identified set $\Theta_0(P_0)$ which we characterize by its support function. Since the identified set depends on the distribution P_0 of X_i so does the corresponding support function. We capture such dependence by the composition mapping :

$$\nu(p, \Theta_0(Q)) = \sup_{\theta \in \Theta_0(Q)} \langle p, \theta \rangle. \quad (17)$$

⁵For a detailed geometric discussion of $p \mapsto \nu(p, K)$, see Beresteanu and Molinari (2008) and Kaido (2010).

As P_0 is unknown, we view $\nu(\cdot, \Theta_0(P_0))$ as an infinite dimensional parameter defined on $\mathcal{C}(\mathbb{S}^{d_\theta})$ and aim to characterize its semiparametric efficiency bound.

3.1.2 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 et al. (1993) for a full discussion.

Let \mathbf{M} denote the set of Borel probability measures on \mathcal{X} , endowed with the τ -topology,⁶ and μ be a positive σ -finite measure such that P_0 is absolutely continuous with respect to μ (denoted $P_0 \ll \mu$). Of particular interest is the set $\mathbf{M}_\mu \equiv \{P \in \mathbf{M} : P \ll \mu\}$, which consists of the measures in \mathbf{M} that are absolutely continuous with respect to μ . This set may be identified with the space:

$$L_\mu^2 \equiv \{f : \mathcal{X} \rightarrow \mathbf{R} : \|f\|_{L_\mu^2} < \infty\} \quad \|f\|_{L_\mu^2}^2 \equiv \int f^2(x) d\mu(x) \quad (18)$$

via the mapping $P \mapsto \sqrt{dP/d\mu}$. A model $\mathbf{P} \subseteq \mathbf{M}_\mu$ is then just a collection of probability measures which, with some abuse of notation, we may also consider as a subset of L_μ^2 .

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

Definition 3.1. A function $h : N \rightarrow L_\mu^2$ is a curve in L_μ^2 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}_η denote its Fréchet derivative at any point $\eta \in N$. ■

Definition 3.2. For $\mathbf{P} \subseteq L_\mu^2$ and a function $p_0 \in \mathbf{P}$, the tangent set of \mathbf{P} at p_0 is defined as:

$$\dot{\mathbf{P}}^0 \equiv \{\dot{h}_\eta : \eta \mapsto h_\eta \text{ is a curve in } L_\mu^2 \text{ with } h_0 = p_0 \text{ and } h_\eta \in \mathbf{P} \text{ for all } \eta\} . \quad (19)$$

The tangent space of \mathbf{P} at p_0 , denoted by $\dot{\mathbf{P}}$, is the closure of the linear span of $\dot{\mathbf{P}}^0$ (in L_μ^2). ■

A curve $\eta \mapsto h_\eta$ with $h_\eta \in \mathbf{P}$ for all η is therefore a parametric submodel with the condition that it be quadratic mean differentiable being equivalent to $\eta \mapsto h_\eta$ being Fréchet differentiable. In turn, the tangent set at P_0 is the subspace of L_μ^2 spanned by the scores corresponding to the parametric submodels passing through P_0 . The difference between the efficiency analysis of finite dimensional parameters and the present context, is in the appropriate notion of differentiability for the parameter of interest as a function of the underlying true distribution P_0 .

A parameter defined on $\mathcal{C}(\mathbb{S}^{d_\theta})$ is a mapping $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ that assigns to each probability measure $Q \in \mathbf{P}$ a corresponding function in $\mathcal{C}(\mathbb{S}^{d_\theta})$. In our context, ρ assigns to a measure Q the support function of its identified set; that is $\rho(Q) = \nu(\cdot, \Theta_0(Q))$. In order to derive a semiparametric efficiency bound for estimating $\rho(P_0)$, we require that the mapping $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ be smooth in the sense of being pathwise weak-differentiable.

⁶The 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 $f : \mathcal{X} \rightarrow \mathbf{R}$. Note that unlike the weak topology, continuity of f is not required.

Definition 3.3. For a model $\mathbf{P} \subseteq L_\mu^2$ and a parameter $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ we say ρ is pathwise weak-differentiable at $p_0 \in \mathbf{P}$ if there exists a continuous linear operator $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ such that:

$$\lim_{\eta \rightarrow 0} \left| \int_{\mathbb{S}^{d_\theta}} \left\{ \frac{\rho(h_\eta)(p) - \rho(h_0)(p)}{\eta} - \dot{\rho}(\dot{h}_0)(p) \right\} dB(p) \right| = 0, \quad (20)$$

for any finite Borel measure B on \mathbb{S}^{d_θ} and any curve $\eta \mapsto h_\eta$ with $h_\eta \in \mathbf{P}$ for all η and $h_0 = p_0$.⁷ ■

Given these definitions, we can state a precise notion of semiparametric efficiency for estimating $\rho(P_0)$. While semiparametric efficiency may be characterized in multiple ways, we opt to define it in terms of the convolution theorem; see Theorem 5.2.1 in Bickel et al. (1993).

Theorem 3.1. (Convolution Theorem) *Suppose that (i) $P_0 \in \mathbf{P}$, (ii) $\dot{\mathbf{P}}^0$ is linear and (iii) $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ is pathwise weak-differentiable at P_0 . Then, there exists a tight mean zero Gaussian process \mathbb{G}_0 in $\mathcal{C}(\mathbb{S}^{d_\theta})$ such that if $\{T_n\}$ is a regular estimator for $\rho(P_0)$ converging in law to \mathbb{G} :*

$$\mathbb{G} \stackrel{L}{=} \mathbb{G}_0 + \Delta_0,$$

where $\stackrel{L}{=}$ denotes equality in law and Δ_0 is a tight Borel measurable random element in $\mathcal{C}(\mathbb{S}^{d_\theta})$ independent of \mathbb{G}_0 .⁸

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 aim to obtain the covariance kernel for the Gaussian process \mathbb{G}_0 , denoted:

$$I^{-1}(p_1, p_2) \equiv \text{Cov}(\mathbb{G}_0(p_1), \mathbb{G}_0(p_2)) \quad (21)$$

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

3.2 Efficiency Bound

3.2.1 Assumptions

We characterize the distribution of the efficient limiting process \mathbb{G}_0 under the following assumptions.

Assumption 3.1. (i) $\Theta \subset \mathbf{R}^{d_\theta}$ is convex and compact; (ii) $\Theta_0(P_0)$ is contained in the interior of Θ (relative to \mathbf{R}^{d_θ}); (iii) There exists a $\theta_0 \in \Theta$ such that $F(E[m(X_i, \theta_0)]) < 0$.

⁷The dual space of $\mathcal{C}(\mathbb{S}^{d_\theta})$ is the set of finite Borel measures on \mathbb{S}^{d_θ} . For $\mathbf{B} \equiv \mathcal{C}(\mathbb{S}^{d_\theta})$, $\rho : \mathbf{P} \rightarrow \mathbf{B}$, \mathbf{B}^* the dual of \mathbf{B} and $\langle b, b^* \rangle_{\mathbf{B}} = b^*(b)$ for any $b \in \mathbf{B}$ and $b^* \in \mathbf{B}^*$, Definition 3.3 amounts to $\langle \eta^{-1} \{ \rho(h_\eta) - \rho(h_0) \} - \dot{\rho}(\dot{h}_0), b^* \rangle_{\mathbf{B}} = o(1)$.

⁸ $\{T_n\}$ 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{P} passing through $p_0 \equiv \sqrt{dP_0/d\mu}$ and every $\{\eta_n\}$ with $\eta_n = O(n^{-\frac{1}{2}})$, $\sqrt{n}(T_n - \rho(h_{\eta_n})) \xrightarrow{L_n} \mathbb{G}$ where L_n is the law under $P_{\eta_n}^n$.

Assumption 3.2. (i) $m : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^{d_m}$ 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) $\theta \mapsto \nabla m(x, \theta)$ is equicontinuous in $x \in \mathcal{X}$.

Assumption 3.3. (i) There is $V \subset \mathbf{R}^{d_m}$ with $v \mapsto F(v)$ differentiable on $v \in V$; (ii) $v \mapsto \nabla F(v)$ uniformly continuous and bounded in $v \in V$; (iii) There is a neighborhood $N(P_0) \subseteq \mathbf{M}$ of P_0 such that $\int m(x, \theta) dQ(x) \in V$ for all $(\theta, Q) \in \Theta \times N(P_0)$.

Assumption 3.4. (i) $P_0 \ll \mu$ for a positive σ -finite μ ; (ii) $\theta \mapsto F^{(i)}(\int m_S(x, \theta) dQ(x))$ is strictly convex for all $Q \in N(P_0)$ and $1 \leq i \leq d_{F_S}$; (iii) $\#\mathcal{A}(\theta, P_0) \leq d_{\theta}$ for all $\theta \in \Theta_0(P_0)$; (iv) $\{\nabla F^{(i)}(E[m(X, \theta)])E[\nabla_{\theta} m(X, \theta)]\}_{i \in \mathcal{A}(\theta, P_0)}$ are linearly independent for all $\theta \in \Theta_0(P_0)$.

The requirement that Θ be convex, imposed in Assumption 3.1(i), helps ensure the identified set is convex as well. In turn, Assumption 3.1(ii) implies that the boundary of $\Theta_0(P_0)$ is characterized by the inequality constraints and not by the parameter space Θ . Our Assumptions, however, still allow for norm constraints on θ ; see Remark 3.2 below. Assumption 3.1(iii) imposes that $\Theta_0(P_0)$ have a nonempty interior (relative to $\mathbf{R}^{d_{\theta}}$) and it enables us to represent the support function as the solution to a Lagrangian. Additionally, Assumption 3.1(iii) effectively rules out the presence of moment equalities, though we note strictly convex moment equalities would imply the model is either identified or the identified set is nonconvex. Assumptions 3.2(i)-(ii) require $m(x, \theta)$ and $\nabla_{\theta} m(x, \theta)$ to be bounded in $(x, \theta) \in \mathcal{X} \times \Theta$, which for some parameterizations may imply that $\mathcal{X} \subseteq \mathbf{R}^{d_x}$ be compact. We note that continuity in $x \in \mathcal{X}$ is not imposed in Assumption 3.2, which is the main reason we employ the τ -topology rather than the weak topology on \mathbf{M} . This assumption allows us to incorporate examples with moment restrictions that involve indicator functions.

Assumptions 3.3(i)-(iii) require F to be continuously differentiable on the range of the mapping $(\theta, Q) \mapsto \int m(x, \theta) dQ(x)$ from Θ and a neighborhood of P_0 . This condition is satisfied if F is continuously differentiable on a neighborhood of the set $\{E[m(X_i, \theta)]\}_{\theta \in \Theta}$. Assumption 3.4(ii) yields that $\Theta_0(Q)$ is also convex for any probability measure Q close to P_0 (in the τ -topology). In the moment inequalities model, Assumption 3.4(ii) holds if for all $1 \leq i \leq d_{F_S}$ the functions $\theta \mapsto m_S^{(i)}(x, \theta)$ are weakly convex at all $x \in \mathcal{X}$ and strictly convex on some set $\mathcal{X}_{0,i} \subset \mathcal{X}$ with positive probability under P_0 . Assumption 3.4(iii) imposes that the number of active constraints at each $\theta \in \Theta_0(P_0)$ be less than or equal to d_{θ} . This requirement plays a crucial role in ensuring that the support function is pathwise weak-differentiable. For example, Hirano and Porter (2009) show that a scalar-valued parameter defined as the minimum of population means is not differentiable when multiple means attain the minimum – i.e. when the number of active constraints is larger than the dimension of the parameter. While Assumption 3.4(iii) is restrictive for scalar valued parameters, it is more easily verified in higher dimensional models. Finally, Assumption 3.4(iv) implies the Lagrange multipliers associated with the support function are uniquely defined.

Remark 3.1. Requiring that linear constraints be of the form $A\theta - F_L(E[m_L(X_i)])$ for some known matrix A 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, Θ be a ball of radius B in \mathbf{R}^2 , and denote $x = (x^{(1)}, x^{(2)})'$,

$\theta = (\theta^{(1)}, \theta^{(2)})'$. Suppose F is the identity and that for some $K > 0$, $m : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^3$ is given by:

$$m^{(1)}(x, \theta) \equiv x^{(1)}\theta^{(1)} + x^{(2)}\theta^{(2)} - K \quad m^{(2)}(x, \theta) \equiv -\theta^{(2)} \quad m^{(3)}(x, \theta) \equiv -\theta^{(1)}. \quad (22)$$

It can then be verified Assumptions 3.1-3.4 are satisfied provided $P_0 \ll \mu$, $E[X^{(1)}] > 0$, $E[X^{(2)}] > 0$, and B is sufficiently large. Consider a curve $\eta \mapsto h_\eta$ in L_μ^2 with $h_0 \equiv \sqrt{dP_0/d\mu}$ and satisfying:

$$\int h_\eta^2(x)d\mu(x) = 1 \quad \int x^{(1)}h_\eta^2(x)d\mu(x) = E[X^{(1)}](1 + \eta) \quad \int x^{(2)}h_\eta^2(x)d\mu(x) = E[X^{(2)}], \quad (23)$$

and note that if P_η is such that $h_\eta = \sqrt{dP_\eta/d\mu}$, then P_η satisfies Assumptions 3.1-3.4 with P_η in place of P_0 . However, at the point $\bar{p} \equiv \bar{s}/\|\bar{s}\|$ with $\bar{s} \equiv (E[X^{(1)}], E[X^{(2)}])'$, we obtain that:

$$\nu(\bar{p}, \Theta_0(P_\eta)) = \begin{cases} \frac{K}{\|\bar{s}\|} & \text{if } \eta \geq 0 \\ \frac{K}{\|\bar{s}\|} \frac{E[X^{(1)}]}{(E[X^{(1)}] + \eta)} & \text{if } \eta < 0 \end{cases}, \quad (24)$$

which implies the support function is not pathwise weak-differentiable at $\eta = 0$.⁹ Contrary to this specification, Examples 2.1 and 2.2 involve linear constraints that take the form $A\theta - F_L(E[m_L(X)])$ for a known matrix A and thus can be studied within our framework. ■

Remark 3.2. We note that strictly convex norm constraints, such as $\|\theta\|^2 \leq B$, can be accommodated by setting, for example, $F^{(i)}(\int m(X, \theta)dQ(x)) \equiv \|\theta\|^2 - B$ for some $1 \leq i \leq d_{F_S}$ and all Q . Similarly, upper or lower bound constraints on individual elements $\theta^{(i)}$ of the parameter vector θ may be imposed through the linear constraints $A\theta - F_L(E[m_L(X)]) \leq 0$. ■

3.2.2 Inverse Information Covariance Functional

Since the limiting efficient process \mathbb{G}_0 is mean zero Gaussian, its distribution is fully characterized by its covariance kernel $I^{-1} : \mathbb{S}^{d_\theta} \times \mathbb{S}^{d_\theta} \rightarrow \mathbf{R}$. The semiparametric efficiency bound may therefore be described in terms of the inverse information covariance functional I^{-1} , much in the same way the limiting efficient covariance matrix is the object reported in finite dimensional problems.

Before deriving a closed form for I^{-1} , we require to introduce some additional notation. Since the moment restrictions are convex in θ , the support function:

$$\nu(p, \Theta_0(P_0)) = \sup_{\theta \in \Theta} \{ \langle p, \theta \rangle \text{ s.t. } F(\int m(x, \theta)dP_0(x)) \leq 0 \} \quad (25)$$

is in fact the maximum value of an ordinary convex program (Rockafellar, 1970). Consequently, the support function also admits a Lagrangian representation:

$$\nu(p, \Theta_0(P_0)) = \sup_{\theta \in \Theta} \{ \langle p, \theta \rangle + \lambda(p, P_0)' F(\int m(x, \theta)dP_0(x)) \}, \quad (26)$$

where, under our assumptions, the Lagrange multipliers $\lambda(p, P_0) \in \mathbf{R}^{d_F}$ are unique for all $p \in \mathbb{S}^{d_\theta}$.¹⁰ Moreover, the set of maximizers for both (25) and (26) is the same, consisting of the boundary

⁹We are indebted to Mark Machina for this example.

¹⁰Lemma B.7 in the Appendix in fact establishes this representation holds for all $p \mapsto \nu(p, \Theta_0(Q))$ with $Q \in N(P_0)$.

points of $\Theta_0(P_0)$ at which $\Theta_0(P_0)$ is tangent to the hyperplane $\{\theta \in \Theta : \langle p, \theta \rangle = \nu(p, \Theta_0(P_0))\}$. These boundary points, together with their associated Lagrange multipliers, are instrumental in both characterizing the semiparametric efficiency bound and the efficient estimator.

Theorem 3.2 derives the inverse information covariance functional for estimating $p \mapsto \nu(p, \Theta_0(P_0))$.

Theorem 3.2. *Let Assumptions 3.1-3.4 hold. For each $\theta \in \Theta$, let $H(\theta) \equiv \nabla F(E[m(X_i, \theta)])$, and for each $\theta_1, \theta_2 \in \Theta$, let $\Omega(\theta_1, \theta_2) \equiv E[(m(X_i, \theta_1) - E[m(X_i, \theta_1)])(m(X_i, \theta_2) - E[m(X_i, \theta_2)])']$. Then:*

$$I^{-1}(p_1, p_2) = \lambda(p_1, P_0)' H(\theta^*(p_1)) \Omega(\theta^*(p_1), \theta^*(p_2)) H(\theta^*(p_2))' \lambda(p_2, P_0) , \quad (27)$$

for any $\theta^*(p_1) \in \arg \max_{\theta \in \Theta_0(P_0)} \langle p_1, \theta \rangle$ and any $\theta^*(p_2) \in \arg \max_{\theta \in \Theta_0(P_0)} \langle p_2, \theta \rangle$.

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:

$$\text{Var}\{\lambda(\bar{p}, P_0)' \nabla F(E[m(X_i, \theta^*(\bar{p})]) m(X_i, \theta^*(\bar{p}))\} , \quad (28)$$

for any $\theta^*(\bar{p}) \in \arg \max_{\theta \in \Theta_0(P_0)} \langle \bar{p}, \theta \rangle$. Hence, since Lagrange multipliers corresponding to non-binding moment inequalities are zero, the semiparametric efficiency bound for $\nu(\bar{p}, \Theta_0(P_0))$ is the variance of a linear combination of the binding constraints at the boundary point $\theta^*(\bar{p}) \in \partial\Theta_0(P_0)$. 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} – i.e. in increasing the value of the support function at \bar{p} . Thus, the semiparametric efficiency bound is the variance of a linear combination of binding constraints, where the weight each constraint receives is determined by its importance in shaping the boundary of the identified set at the point $\theta^*(\bar{p}) \in \partial\Theta_0(P_0)$.

Remark 3.3. In the event (26) admits multiple maximizers, the definition of $I^{-1}(p_1, p_2)$ is invariant to the choice of $\theta^*(p_1)$ or $\theta^*(p_2)$. If θ_1^*, θ_2^* maximize (26), then the complementary slackness condition implies only Lagrange multipliers corresponding to linear constraints that are active at both θ_1^* and θ_2^* may be nonzero. Therefore, letting $\lambda_L(p, P_0)$ denote the Lagrange multiplier for linear constraints, we obtain from the corresponding complementary slackness condition that:

$$\lambda(p, P_0)' H(\theta^*) m(x, \theta^*) = \lambda_L(p, P_0)' (F_L(E[m_L(X_i)]) - \nabla F_L(E[m_L(X_i)]) m_L(x)) , \quad (29)$$

for $\theta^* = \theta_1^*$ and $\theta^* = \theta_2^*$. Hence, $I^{-1}(p_1, p_2)$ is independent of the choice of maximizer. ■

4 Efficient Estimation

Employing the characterization of the semiparametric efficiency bound obtained in Theorem 3.2, this section establishes that the intuitive “plug-in” estimator is in fact efficient. A bootstrap procedure is also shown to consistently estimate the distribution of \mathbb{G}_0 in $\mathcal{C}(\mathbb{S}^{d_\theta})$.

4.1 The Estimator

We impose the additional requirement that the data be independent and identically distributed.

Assumption 4.1. $\{X_i\}_{i=1}^n$ is an i.i.d. sample with each X_i distributed according to P_0 .

Given a sample $\{X_i\}_{i=1}^n$, we let \hat{P}_n denote the empirical measure – that is \hat{P}_n is the random discrete measure assigning to each $x \in \mathcal{X}$ probability $\hat{P}_n(x) \equiv \frac{1}{n} \sum_i 1\{X_i = x\}$. Under Assumption 4.1, \hat{P}_n is consistent for the unknown distribution P_0 under the τ -topology. Therefore, a natural estimator for the support function $p \mapsto \nu(p, \Theta_0(P_0))$ is its sample analogue pointwise defined by:

$$\nu(p, \Theta_0(\hat{P}_n)) = \sup_{\theta \in \Theta} \{ \langle p, \theta \rangle \text{ s.t. } F(\frac{1}{n} \sum_{i=1}^n m(X_i, \theta)) \leq 0 \} . \quad (30)$$

It is useful to note that in lieu of Assumption 3.4(ii) and the consistency of \hat{P}_n , the constraints in (30) are convex in $\theta \in \Theta$ with probability tending to one. As a result, $p \mapsto \nu(p, \Theta_0(\hat{P}_n))$ also admits a characterization as the maximum of a set of Lagrangians indexed by $p \in \mathbb{S}^{d_\theta}$:

$$\nu(p, \Theta_0(\hat{P}_n)) = \sup_{\theta \in \Theta} \{ \langle p, \theta \rangle + \lambda(p, \hat{P}_n)' F(\frac{1}{n} \sum_{i=1}^n m(X_i, \theta)) \} , \quad (31)$$

where with probability tending to one $\lambda(p, \hat{P}_n) \in \mathbf{R}^{d_F}$ is unique for all $p \in \mathbb{S}^{d_\theta}$ and $p \mapsto \lambda(p, \hat{P}_n)$ is continuous and bounded on \mathbb{S}^{d_θ} . 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 $\nu(\cdot, \Theta_0(P_0))$ and in this manner characterize its influence function in $\mathcal{C}(\mathbb{S}^{d_\theta})$.

Theorem 4.1 reveals that the resulting influence function is in fact the efficient one, and hence $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ is a semiparametric efficient estimator for $\nu(\cdot, \Theta_0(P_0))$.

Theorem 4.1. *If Assumptions 3.1, 3.2, 3.3, 3.4 and 4.1 hold, then: (i) $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ is a regular estimator for $\nu(\cdot, \Theta_0(P_0))$; (ii) Uniformly in $p \in \mathbb{S}^{d_\theta}$:*

$$\sqrt{n}\{\nu(p, \Theta_0(\hat{P}_n)) - \nu(p, \Theta_0(P_0))\} = \lambda(p, P_0)' \frac{1}{\sqrt{n}} \sum_{i=1}^n H(\theta^*(p)) \{m(X_i, \theta^*(p)) - E[m(X_i, \theta^*(p))]\} + o_p(1) ,$$

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 $\mathcal{C}(\mathbb{S}^{d_\theta})$,

$$\sqrt{n}\{\nu(\cdot, \Theta_0(\hat{P}_n)) - \nu(\cdot, \Theta_0(P_0))\} \xrightarrow{L} \mathbb{G}_0 , \quad (32)$$

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

In moment inequality models, it is common to find that the limiting distribution of statistics $\{T_n(\theta)\}$ depend on the inequalities that bind at each $\theta \in \Theta_0(P_0)$. As a result, the limiting distribution of $\{T_n(\theta)\}$ is often discontinuous in θ , even if we restrict attention to the boundary of the identified set $\partial\Theta_0(P_0)$. It is interesting to note that this type of discontinuity is not present when we focus on support functions instead. In particular, Theorem 4.1 implies all sample paths of the process $p \mapsto \mathbb{G}_0(p)$ are uniformly continuous on \mathbb{S}^{d_θ} . This continuity in $p \in \mathbb{S}^{d_\theta}$ results from the

presence of the Lagrange multipliers in the influence function. Heuristically, $\lambda(p, P_0)$ determines the weight a binding constraint receives at each point $p \in \mathbb{S}^{d_\theta}$. The complementarity slackness condition and continuity of $p \mapsto \lambda(p, P_0)$ then imply that if p_1 and p_2 are close, then constraints that are binding at p_1 but not p_2 must have a correspondingly small Lagrange multipliers. In this manner, $p \mapsto \lambda(p, P_0)$ “smooths” out the empirical process and delivers asymptotic uniform equicontinuity.

Remark 4.1. Despite its simplicity and efficiency, the “plug-in” support function estimator has not been studied in general moment inequalities problems. In related work, Beresteanu and Molinari (2008) employ “plug-in” estimators for identified sets that can be expressed as the Aumann expectation of a set valued random variable. Unfortunately, the models they consider only intersect with ours in very special cases.¹¹ As a result, while we conjecture their estimator is semiparametrically efficient, such a claim does not follow from our results. ■

Remark 4.2. In general, $\nu(\cdot, \hat{\Theta}_0(P_n))$ is not well defined when no $\theta \in \Theta$ satisfies the constraints in (30). Even though the probability of such event vanishes asymptotically, it may occur with non-negligible probability in finite samples when the parameter is near point identification and/or the number of constraints is large. In these instances, we may instead define:

$$\tilde{\nu}_n(p, \Theta_0(\hat{P}_n)) \equiv \sup_{\theta \in \Theta} \{ \langle p, \theta \rangle \text{ s.t. } F\left(\frac{1}{n} \sum_{i=1}^n m(X_i, \theta)\right) \leq \epsilon_n \} , \quad (33)$$

where $0 \leq \epsilon_n \in \mathbf{R}^{d_F}$ and $\|\epsilon_n\| = o_p(n^{-\frac{1}{2}})$. It can be shown that the “slackness” in the constraints ensures $\tilde{\nu}_n(\cdot, \Theta_0(\hat{P}_n))$ is not only well defined with probability one, but also asymptotically equivalent to $\nu(\cdot, \Theta_0(\hat{P}_n))$. As a result, $\tilde{\nu}_n(\cdot, \Theta_0(\hat{P}_n))$ inherits the first order properties of $\nu(\cdot, \Theta_0(\hat{P}_n))$ and, in particular, is also a semiparametric efficient estimator of $\nu(\cdot, \Theta_0(P_0))$. ■

4.1.1 An Estimator for $\Theta_0(P_0)$

Due to the close link between convex compact sets and their support functions, it is straightforward to employ the estimator $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ for $\nu(\cdot, \Theta_0(P_0))$ to obtain a corresponding estimator $\{\hat{\Theta}_n\}$ for $\Theta_0(P_0)$. Specifically, for a set C let $co\{C\}$ denote its convex hull and define:

$$\hat{\Theta}_n \equiv co\{\Theta_0(\hat{P}_n)\} . \quad (34)$$

By Assumption 3.4(ii), the sets $\{\Theta_0(\hat{P}_n)\}$ and $\{\hat{\Theta}_n\}$ are in fact the same with probability tending to one and their support function is just $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$.

The semiparametric efficiency of $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ and the equality of the Hausdorff distance between convex sets and the supremum distance between their corresponding support functions (see (16)) suggests that, in expectation, $\{\hat{\Theta}_n\}$ should be closer to $\Theta_0(P_0)$ than competing estimators. Theorem 4.2 establishes this result, showing that $\{\hat{\Theta}_n\}$ minimizes the asymptotic risk for a wide class of loss functions based on the Hausdorff distance between $\Theta_0(P_0)$ and its estimator.

¹¹Specifically, $\Theta_0(P_0)$ takes the form of an Aumann expectation when $\Theta_0(P_0) = \{\theta \in \Theta : E[L] \leq \theta \leq E[U]\}$ for two random variables L, U ; see also Remark 4.5.

Theorem 4.2. *Let Assumption 3.1, 3.2, 3.3, 3.4 and 4.1 hold and $L : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ be a nondecreasing continuous function satisfying $L(0) = 0$ and $L(a) \leq Ma^\kappa$ for some M, κ and all $a \in \mathbf{R}_+$. If $\{K_n\}$ is a regular convex compact valued set estimator for $\Theta_0(P_0)$, then it follows that:¹²*

$$\liminf_{n \rightarrow \infty} E[L(\sqrt{nd}_H(K_n, \Theta_0(P_0)))] \geq \limsup_{n \rightarrow \infty} E[L(\sqrt{nd}_H(\hat{\Theta}_n, \Theta_0(P_0)))] = E[L(\|\mathbb{G}_0\|_\infty)] . \quad (35)$$

The lower bound on asymptotic risk obtained in (35) is a direct consequence of the Convolution Theorem and in fact holds for any nondecreasing loss function $L : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ with $L(0) = 0$. The continuity of $a \mapsto L(a)$ and that it be majorized by a polynomial are imposed in order to show $\{\hat{\Theta}_n\}$ actually attains the bound. The latter requirement may be weakened to $L(a) \leq M \exp(a\kappa)$ for some $M > 0$ but only for a limited range of values of κ . Furthermore, we also note that (35) also holds if the Hausdorff distance is replaced by any weaker or equivalent metric.¹³

Remark 4.3. The consistency of $\{\hat{\Theta}_n\}$ and its rate of convergence in moment inequalities models has previously been noted in Chernozhukov et al. (2007). The requirements that $\Theta_0(P_0)$ be convex and have non-empty interior (relative to \mathbf{R}^{d_θ}) imply $\Theta_0(P_0)$ may be well approximated from its interior. Such a property was termed a “degeneracy condition” by Chernozhukov et al. (2007), who show it suffices for the plug-in estimator to be consistent. The optimality of such estimator, however, had previously not been shown in the literature. ■

Remark 4.4. Employing the equivalence in (16) and Theorem 4.1, it is straightforward to obtain:

$$\sqrt{nd}_H(\hat{\Theta}_n, \Theta_0(P_0)) \xrightarrow{L} \|\mathbb{G}_0\|_\infty . \quad (36)$$

In Section 4.2 we derive a bootstrap procedure for computing quantiles of limiting distributions as in (36). These results will enable us to obtain confidence intervals using the efficient estimator. ■

Remark 4.5. In settings where θ is scalar and the identified region is the set of $\theta \in \Theta$ satisfying:

$$E[L(X_i)] \leq \theta \leq E[U(X_i)] , \quad (37)$$

for some functions $L : \mathcal{X} \rightarrow \mathbf{R}$ and $U : \mathcal{X} \rightarrow \mathbf{R}$ satisfying $E[L(X_i)] < E[U(X_i)]$, we obtain that $\hat{\Theta}_n = [\frac{1}{n} \sum_i L(X_i), \frac{1}{n} \sum_i U(X_i)]$. Hence, in this special case our estimator reduces to the one examined in Imbens and Manski (2004) and Beresteanu and Molinari (2008), which we believe to be the natural estimator for the problem. ■

4.2 A Consistent Bootstrap

We obtain a consistent bootstrap for estimating the distribution of \mathbb{G}_0 following a “score based” approach proposed in Kline and Santos (2010). In particular, for W_i a mean zero scalar random

¹²We say $\{K_n\}$ is a regular estimator of Θ_0 if its support function $\nu(\cdot, K_n)$ is a regular estimator for $\nu(\cdot, \Theta_0(P_0))$.

¹³For example, for a positive Borel measure B on \mathbb{S}^{d_θ} , the distance $d_p(K_1, K_2) \equiv \|\nu(\cdot, K_1) - \nu(\cdot, K_2)\|_{L_p^B} = (\int_{\mathbb{S}^{d_\theta}} |\nu(q, K_1) - \nu(q, K_2)|^p dB(q))^{1/p}$ can be used. See Schneider (1993) for additional discussion on distance measures for convex sets.

variable and $\{W_i\}_{i=1}^n$ an i.i.d. sample independent of $\{X_i\}_{i=1}^n$, we define:

$$G_n^*(p) \equiv \lambda(p, \hat{P}_n)' \nabla F \left(\frac{1}{n} \sum_{i=1}^n m(X_i, \hat{\theta}(p)) \right) \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ m(X_i, \hat{\theta}(p)) - \frac{1}{n} \sum_{i=1}^n m(X_i, \hat{\theta}(p)) \right\} W_i, \quad (38)$$

where $\lambda(p, \hat{P}_n)$ is as in (31) and $\hat{\theta}(p)$ is any maximizer for the optimization problem in (31). 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 $\{W_i\}_{i=1}^n$.

The perturbation weights $\{W_i\}_{i=1}^n$ are required to satisfy the following weak assumption:

Assumption 4.2. (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 the limiting process \mathbb{G}_0 . Formally, letting L^* denote a law statement conditional on $\{X_i\}_{i=1}^n$, Theorem 4.3 establishes consistency of the law of G_n^* under L^* for that of \mathbb{G}_0 as a stochastic process in $\mathcal{C}(\mathbb{S}^{d_\theta})$.

Theorem 4.3. *If Assumptions 3.1, 3.2, 3.3, 3.4, 4.1 and 4.2 hold, then $G_n^* \xrightarrow{L^*} \mathbb{G}_0$ (in probability).*

In practice, the distribution of G_n^* under L^* can be obtained through simulations of $\{W_i\}_{i=1}^n$. It is also worth noting that the Lagrange multipliers $\lambda(p, \hat{P}_n)$ and maximizers $\hat{\theta}(p)$ are by-products of constructing the full sample estimator of the support function $\nu(\cdot, \Theta_0(\hat{P}_n))$. Therefore, simulating the distribution of G_n^* under L^* only requires sampling perturbation weights $\{W_i\}_{i=1}^n$. As a result, the proposed bootstrap possesses an important computational advantage over the nonparametric bootstrap or the weighted bootstrap (Barbe and Bertail, 1995), which require recomputing the support function in each bootstrap iteration. The procedure even offers advantages over simulating paths of \mathbb{G}_0 from an estimated covariance kernel, since we need only draw perturbation weights and hence an estimate of the covariance kernel is not needed.

4.2.1 Estimating Critical Values

In order to employ the efficient estimator for inference, it is often necessary to estimate quantiles of transformations of the limiting process \mathbb{G}_0 . In this section, we provide a general consistent procedure that applies to all the inference methods we subsequently examine. Specifically, we develop consistent estimators for quantiles of random variables that are of the form:

$$\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p)), \quad (39)$$

where $\Psi_0 \subseteq \mathbb{S}^{d_\theta}$ and $\Upsilon : \mathbf{R} \rightarrow \mathbf{R}$ 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 4.3. (i) $\Upsilon : \mathbf{R} \rightarrow \mathbf{R}$ is continuous; (ii) $\{\hat{\Psi}_n\}$ does not depend on $\{W_i\}_{i=1}^n$ and $\hat{\Psi}_n \subseteq \mathbb{S}^{d_\theta}$ 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 (39) may then be estimated through 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 (31). ■

STEP 2: Generate a random sample $\{W_i\}_{i=1}^n$ satisfying Assumption 4.2 to construct G_n^* . ■

STEP 3: Employing G_n^* and $\{\hat{\Psi}_n\}$, estimate the $1 - \alpha$ quantile of $\sup_{p \in \Psi_0} \Upsilon(G_n(p))$ by:

$$\hat{c}_{1-\alpha} \equiv \inf\{c : P(\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) \geq c \mid \{X_i\}_{i=1}^n) \geq 1 - \alpha\}. \quad (40)$$

In practice, $\hat{c}_{1-\alpha}$ is often not explicitly computable but obtainable through simulation. In particular, by repeating Steps 1 and 2 we may obtain a simulated sample of $\{\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p))\}$. The $1 - \alpha$ sample quantile across simulations then provides an approximation to $\hat{c}_{1-\alpha}$. ■

As Theorem 4.4 establishes, $\hat{c}_{1-\alpha}$ is indeed consistent for the desired quantile.

Theorem 4.4. *Let Assumptions 3.1, 3.2, 3.3, 3.4, 4.1, 4.2 and 4.3 hold and the cdf of $\sup_{p \in \Psi_0} \Upsilon(G_0(p))$ be continuous and strictly increasing at its $1 - \alpha$ quantile, denoted $c_{1-\alpha}$. It then follows that:*

$$\hat{c}_{1-\alpha} \xrightarrow{P} c_{1-\alpha}.$$

5 Confidence Regions

In this section, we employ the efficient estimator to obtain confidence regions for the identified set as well as for the true parameter (Imbens and Manski, 2004). Our constructions exploit links between hypotheses concerning the identified set and hypotheses regarding its support function that have been previously noted in Beresteanu and Molinari (2008), Bontemps et al. (2007) and Kaido (2010).

5.1 Identified Set: One Sided

A commonly imposed requirement for a confidence region of the identified set is that it asymptotically cover $\Theta_0(P_0)$ with a specified probability (Chernozhukov et al., 2007; Romano and Shaikh, 2010). Specifically, for a level of confidence $1 - \alpha$, a confidence region \mathcal{C}_n should satisfy:

$$\liminf_{n \rightarrow \infty} P(\Theta_0(P_0) \subseteq \mathcal{C}_n) \geq 1 - \alpha. \quad (41)$$

In analogy with the analysis of Imbens and Manski (2004) for interval valued parameters, we construct a confidence region \mathcal{C}_n simply by expanding the boundary of the optimal estimator $\hat{\Theta}_n$

(as in (34)) by an appropriate amount. Towards this end, for any scalar $\epsilon > 0$ we define:

$$\hat{\Theta}_n^\epsilon \equiv \{\theta \in \mathbf{R}^{d_\theta} : \inf_{\tilde{\theta} \in \hat{\Theta}_n} \|\theta - \tilde{\theta}\| \leq \epsilon\} . \quad (42)$$

Exploiting the relation between convex sets and their support functions, it is then straightforward to show that a confidence region of the form $\mathcal{C}_n = \hat{\Theta}_n^{c/\sqrt{n}}$ satisfies the desired coverage if c equals:

$$c_{1-\alpha}^{(1)} \equiv \inf\{c : P(\sup_{p \in \mathbb{S}^{d_\theta}} |-\mathbb{G}_0(p)|_+ \geq c) \geq 1 - \alpha\} , \quad (43)$$

where $|a|_+ = \max\{a, 0\}$.¹⁴ While computation of $c_{1-\alpha}^{(1)}$ is unfeasible, the setting is a special case of Theorem 4.4 with $\Upsilon(a) = | - a|_+$ and $\Psi_0 = \mathbb{S}^{d_\theta}$. Therefore, a consistent estimator is given by:

$$\hat{c}_{1-\alpha}^{(1)} \equiv \inf\{c : P(\sup_{p \in \mathbb{S}^{d_\theta}} | - G_n^*(p)|_+ \geq c \mid \{X_i\}_{i=1}^n) \geq 1 - \alpha\} . \quad (44)$$

The following Theorem establishes that expanding the estimator $\hat{\Theta}_n$ by a factor of $\hat{c}_{1-\alpha}^{(1)}/\sqrt{n}$ indeed results in a confidence region with the desired coverage property.

Theorem 5.1. *Suppose Assumptions 3.1, 3.2, 3.3, 3.4, 4.1 and 4.2 hold and let $\alpha \in (0, 0.5)$. If in addition there exists a $p_0 \in \mathbb{S}^{d_\theta}$ such that $\text{Var}\{\mathbb{G}_0(p_0)\} > 0$, then it follows that:*

$$\lim_{n \rightarrow \infty} P(\Theta_0(P_0) \subseteq \hat{\Theta}_n^{\hat{c}_{1-\alpha}^{(1)}/\sqrt{n}}) = 1 - \alpha . \quad (45)$$

It is useful to note that we may recover the proposed confidence region from its support function $p \mapsto \{\nu(p, \Theta_0(\hat{P}_n)) + \hat{c}_{1-\alpha}^{(1)}/\sqrt{n}\}$. As a result, computing $\hat{\Theta}_n$ or its expansion by a factor of $\hat{c}_{1-\alpha}^{(1)}/\sqrt{n}$ is unnecessary. Instead, the proposed confidence region may be obtained from the efficient estimator for the support function $p \mapsto \nu(p, \Theta_0(\hat{P}_n))$ and the bootstrap critical value $\hat{c}_{1-\alpha}^{(1)}$.

Remark 5.1. The constructed confidence region has a natural interpretation as the inversion of a hypothesis test. In particular, for any convex compact set $K \subseteq \Theta$, consider testing the null hypothesis $K \subseteq \Theta_0(P_0)$ against the alternative $K \not\subseteq \Theta_0(P_0)$ through the test statistic:

$$T_n(K) \equiv \sqrt{n} \vec{d}_H(K, \hat{\Theta}_n) . \quad (46)$$

Results in Beresteanu and Molinari (2008) then imply $\hat{c}_{1-\alpha}^{(1)}$ is an asymptotically valid critical value for $T_n(K)$ and $\hat{\Theta}_n^{\hat{c}_{1-\alpha}^{(1)}/\sqrt{n}}$ is the largest convex set K for which we fail to reject the null hypothesis. ■

5.2 Identified Set: Two Sided

The results from Section 4 may also be employed to obtain a two sided confidence region for $\Theta_0(P_0)$. Such a confidence region consists of two sets $(\mathcal{C}_n^I, \mathcal{C}_n^O)$ satisfying the coverage requirement:

$$\liminf_{n \rightarrow \infty} P(\mathcal{C}_n^I \subseteq \Theta_0(P_0) \subseteq \mathcal{C}_n^O) \geq 1 - \alpha , \quad (47)$$

¹⁴For conciseness, we relegate details of why $c_{1-\alpha}^{(1)}$ delivers size control to the Appendix. See also Theorem 2.4 in Beresteanu and Molinari (2008) for an application of this construction to iid samples of set valued random variables.

for a desired confidence level of $1 - \alpha$.¹⁵ Under this construction, the boundaries of \mathcal{C}_n^I and \mathcal{C}_n^O provide lower and upper bounds for that of $\Theta_0(P_0)$. Heuristically, the confidence region $(\mathcal{C}_n^I, \mathcal{C}_n^O)$ may therefore be interpreted as a uniform confidence interval for the boundary of the identified set.

A natural way in which to construct the desired confidence region is to expand and contract $\hat{\Theta}_n$ to obtain \mathcal{C}_n^O and \mathcal{C}_n^I respectively. Formally, for any $\epsilon > 0$ we define the ϵ -contraction of $\hat{\Theta}_n$ by:

$$\hat{\Theta}_n^{-\epsilon} \equiv \{\theta \in \hat{\Theta}_n : \inf_{\tilde{\theta} \in \mathbf{R}^{d_\theta} \setminus \hat{\Theta}_n} \|\theta - \tilde{\theta}\| \geq \epsilon\}, \quad (48)$$

and observe $\hat{\Theta}_n^{-\epsilon}$ may potentially be empty for ϵ sufficiently large. Two sided confidence regions of the form $(\hat{\Theta}_n^{-c/\sqrt{n}}, \hat{\Theta}_n^{c/\sqrt{n}})$ then satisfy (47) provided c is set to equal:

$$c_{1-\alpha}^{(2)} \equiv \inf\{c : P(\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)| \geq c) \geq 1 - \alpha\}. \quad (49)$$

Moreover, estimation of the unfeasible critical value $c_{1-\alpha}^{(2)}$ may be readily accomplished employing the bootstrap procedure derived in Section 4.2.1. In particular, setting $\Upsilon(a) = |a|$ and $\Psi_0 = \hat{\Psi}_n = \mathbb{S}^{d_\theta}$ in Theorem 4.4 implies a consistent estimator for $c_{1-\alpha}^{(2)}$ is given by:

$$\hat{c}_{1-\alpha}^{(2)} \equiv \inf\{c : P(\sup_{p \in \mathbb{S}^{d_\theta}} |G_n^*(p)| \geq c \mid \{X_i\}_{i=1}^n) \geq 1 - \alpha\}. \quad (50)$$

The following Theorem verifies the proposed confidence region provides the desired coverage.

Theorem 5.2. *Let Assumptions 3.1, 3.2, 3.3, 3.4, 4.1 and 4.2 hold and let $\alpha \in (0, 1)$. If in addition there exists a $p_0 \in \mathbb{S}^{d_\theta}$ such that $\text{Var}\{\mathbb{G}_0(p_0)\} > 0$, then it follows that:*

$$\lim_{n \rightarrow \infty} P(\hat{\Theta}_n^{-\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}} \subseteq \Theta_0(P_0) \subseteq \hat{\Theta}_n^{\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}}) = 1 - \alpha. \quad (51)$$

Provided the inner set of the confidence region is nonempty, its support function is given by $p \mapsto \{\nu(p, \Theta_0(\hat{P}_n)) - \hat{c}_{1-\alpha}^{(2)}/\sqrt{n}\}$. Therefore, as in Theorem 5.1, it is easier to recover $\hat{\Theta}_n^{-\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}}$ from its support function than to directly compute the contraction of the estimator $\hat{\Theta}_n$.

Remark 5.2. The proposed confidence region can also be obtained by inverting a test of the null hypothesis $K = \Theta_0(P_0)$ against the alternative $K \neq \Theta_0(P_0)$ employing the test statistic:

$$T_n(K) \equiv \sqrt{n}d_H(K, \hat{\Theta}_n) \quad (52)$$

and critical value $\hat{c}_{1-\alpha}^{(2)}$. The regions $\hat{\Theta}_n^{-\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}}$ and $\hat{\Theta}_n^{\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}}$ are then respectively the intersection and union of all compact convex sets K for which we fail to reject the null hypothesis. ■

5.3 Parameter Confidence Regions

A third possible type of confidence region for $\Theta_0(P_0)$ covers each parameter in the identified set with pre-specified probability, as originally proposed in Imbens and Manski (2004). Formally, for a

¹⁵To the best of our knowledge, the two sided confidence region for the identified set hasn't been developed elsewhere. We are indebted to Francesca Molinari for proposing that we investigate this type of coverage requirement.

desired confidence level $1 - \alpha$, we require the confidence region \mathcal{P}_n to satisfy:

$$\inf_{\theta \in \Theta_0(P_0)} \liminf_{n \rightarrow \infty} P(\theta \in \mathcal{P}_n) \geq 1 - \alpha . \quad (53)$$

A confidence region with the desired coverage property can be obtained by inverting a test of the null hypothesis that $\theta \in \Theta_0(P_0)$ against the alternative $\theta \notin \Theta_0(P_0)$ (Romano and Shaikh, 2008). In a similar setting to ours, Kaido (2010) proposes inverting the test statistic:¹⁶

$$H_n(\theta) \equiv \sqrt{n} \vec{d}_H(\{\theta\}, \hat{\Theta}_n) , \quad (54)$$

which converges to a tight random variable if $\theta \in \Theta_0(P_0)$ and diverges to infinity otherwise. The appropriate critical value for the test statistic $H_n(\theta)$ then depends on $\theta \in \Theta$ and is given by:

$$c_{1-\alpha}(\theta) \equiv \inf\{c : P(\sup_{p \in \mathfrak{M}(\theta)} | -\mathbb{G}_0(p)|_+ \geq c) \geq 1 - \alpha\} , \quad (55)$$

where $\mathfrak{M}(\theta)$ is the set of maximizers of the function $p \mapsto \{\nu(p, \{\theta\}) - \nu(p, \Theta_0(P_0))\}$ on \mathbb{S}^{d_θ} .

Estimating the critical value $c_{1-\alpha}(\theta)$ requires us to obtain a consistent estimator for the set $\mathfrak{M}(\theta) \subseteq \mathbb{S}^{d_\theta}$. Towards this end, we follow Kaido (2010) and define the set estimator:

$$\hat{\mathfrak{M}}_n(\theta) \equiv \{p \in \mathbb{S}^{d_\theta} : \{\nu(p, \{\theta\}) - \nu(p, \Theta_0(\hat{P}_n))\} \geq \sup_{\tilde{p} \in \mathbb{S}^{d_\theta}} \{\nu(\tilde{p}, \{\theta\}) - \nu(\tilde{p}, \Theta_0(\hat{P}_n))\} - \frac{\kappa_n}{\sqrt{n}}\} , \quad (56)$$

which satisfies $d_H(\mathfrak{M}(\theta), \hat{\mathfrak{M}}_n(\theta)) = o_p(1)$ provided $\kappa_n = o(n^{\frac{1}{2}})$ and $\kappa_n \uparrow \infty$. Applying Theorem 4.4 with $\Upsilon(a) = |a|_+$, $\Psi_0 = \mathfrak{M}(\theta)$ and $\hat{\Psi}_n = \hat{\mathfrak{M}}_n(\theta)$ then implies a consistent estimate of $c_{1-\alpha}(\theta)$ is:

$$\hat{c}_{1-\alpha}(\theta) \equiv \inf\{c : P(\sup_{p \in \hat{\mathfrak{M}}_n(\theta)} | -G_n^*(p)|_+ \geq c \mid \{X_i\}_{i=1}^n) \geq 1 - \alpha\} . \quad (57)$$

As Theorem 5.3 shows, a confidence region built by inverting a test that rejects the null hypothesis $\theta \in \Theta_0(P_0)$ whenever $H_n(\theta) > \hat{c}_{1-\alpha}(\theta)$ indeed satisfies the coverage requirement in (53).

Theorem 5.3. *Let Assumptions 3.1, 3.2, 3.3, 3.4, 4.1 and 4.2 hold, $\alpha \in (0, 0.5)$ and $\kappa_n \uparrow \infty$ with $\kappa_n = o(n^{\frac{1}{2}})$. If $\text{Var}\{\mathbb{G}_0(p)\} > 0$ for all $p \in \mathbb{S}^{d_\theta}$ and $\hat{\mathcal{P}}_n \equiv \{\theta \in \Theta : H_n(\theta) \leq \hat{c}_{1-\alpha}(\theta)\}$, then:*

$$\inf_{\theta \in \Theta_0(P_0)} \liminf_{n \rightarrow \infty} P(\theta \in \hat{\mathcal{P}}_n) \geq 1 - \alpha . \quad (58)$$

5.3.1 Local Properties

Because the statistic $H_n(\theta)$ is based on the semiparametric efficient estimator, the test that rejects whenever $H_n(\theta) > \hat{c}_{1-\alpha}(\theta)$ satisfies a local optimality property for testing the hypothesis:

$$H_0 : \theta \in \Theta_0(P_0) \quad H_1 : \theta \notin \Theta_0(P_0) . \quad (59)$$

¹⁶The analysis in Kaido (2010) is based on an arbitrary estimator of $\nu(\cdot, \Theta_0(P_0))$, not necessarily the efficient one. This type of test statistic was first studied using a support function approach by Bontemps et al. (2007) for a different class of partially identified models.

Specifically, we show the power function of any test that can control size over local parametric submodels must be everywhere weakly smaller than that of a test based on $H_n(\theta)$ for all $\theta \in \partial\Theta_0(P_0)$ that are supported by a unique hyperplane. Therefore, the proposed test is locally asymptotically uniformly most powerful for such $\theta \in \partial\Theta_0(P_0)$ in the sense of van der Vaart (1999).

A formal statement of the optimality result first requires us to introduce additional notation. Let a curve $\eta \mapsto h_\eta$ in L_μ^2 be a “submodel” if for each η there is a probability measure P_η with $h_\eta = \sqrt{dP_\eta/d\mu}$ and P_η satisfying Assumptions 3.1(ii)-(iii), 3.3(iii) and 3.4(i)-(iv) with P_η in place of P_0 . For any point $\theta \in \partial\Theta_0(P_0)$, then let $\mathbf{H}(\theta)$ denote the set of submodels that satisfy:

$$(i) h_0 = \sqrt{dP_0/d\mu} \quad (ii) \theta \in \Theta_0^o(P_\eta) \text{ if } \eta < 0 \quad (iii) \theta \notin \Theta_0(P_\eta) \text{ if } \eta > 0. \quad (60)$$

Thus, $\mathbf{H}(\theta)$ is the set of submodels passing through P_0 for which the identified set changes from including θ in its interior to not having θ as an element of it. Finally, we denote power functions as mappings $\pi : \mathbf{H}(\theta) \rightarrow [0, 1]$, where $\pi(h_\eta)$ is to be interpreted as the probability that a (possibly randomized) test rejects the null hypothesis in (59) when $X_i \sim P_\eta$.

Given the introduced notation, we formally establish the local optimality of the proposed test.

Theorem 5.4. *Let Assumptions 3.1, 3.2, 3.3, 3.4, 4.1 and 4.2 hold and $\theta_0 \in \partial\Theta_0(P_0)$ with $\mathfrak{M}(\theta_0) = \{p_0\}$ and $\text{Var}\{\mathbb{G}_0(p_0)\} > 0$. Suppose $\{\pi_n\}$ is a sequence of power functions such that:*

$$\limsup_{n \rightarrow \infty} \pi_n(h_{\eta/\sqrt{n}}) \leq \alpha \quad (61)$$

for every $h_\eta \in \mathbf{H}(\theta_0)$ and $\eta \leq 0$. If $\{\pi_n^*\}$ is the power function of the test that rejects when $H_n(\theta_0) > \hat{c}_{1-\alpha}(\theta_0)$, then $\{\pi_n^*\}$ satisfies (61) and for $\tilde{l}(x) = -\lambda(p_0, P_0)'H(\theta_0)\{m(x, \theta_0) - E[m(X_i, \theta_0)]\}$:

$$\limsup_{n \rightarrow \infty} \pi_n(h_{\eta/\sqrt{n}}) \leq \lim_{n \rightarrow \infty} \pi_n^*(h_{\eta/\sqrt{n}}) = 1 - \Phi\left(z_{1-\alpha} - \eta \frac{2E[\tilde{l}(X_i)\dot{h}_0(X_i)/h_0(X_i)]}{\sqrt{E[\mathbb{G}_0^2(p_0)]}}\right), \quad (62)$$

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

Theorem 5.4 applies to tests of (59) at points $\theta \in \partial\Theta_0(P_0)$ for which $\mathfrak{M}(\theta)$ is a singleton. Such $\theta \in \partial\Theta_0(P_0)$ can be thought of as lying in a “smooth” part of the boundary of $\Theta_0(P_0)$ in the sense of being supported by a unique hyperplane. At such a point, the support functions $\nu(\cdot, \{\theta\})$ and $\nu(\cdot, \Theta_0(P_0))$ are tangent at a unique normal vector p_0 and the properties of tests of (59) may be related to those of one sided tests for the parameter $\nu(p_0, \{\theta\}) - \nu(p_0, \Theta_0(P_0))$. This connection enables us establish the result by appealing to the local optimality of semiparametric efficient estimators for testing one sided hypothesis on scalar parameters.

It is important to note, however, that as in most of the literature on semiparametric efficiency Theorem 5.4 is local in nature. In particular, the size control requirement in (61) is local and the proposed test does not necessarily control size uniformly over a larger set of distributions. Of special interest would be to generalize the results in Imbens and Manski (2004) and Stoye (2009) to obtain confidence regions based on the efficient estimator that are uniformly valid regardless of $\Theta_0(P_0)$. Such an extension, however, is beyond the scope of the present paper.

6 Simulation Evidence

In this section, we assess the finite sample performance of the efficient estimator and illustrate its ease of implementation with a Monte Carlo experiment based on Example 2.1.

We let $Z_i \equiv (Z_{1,i}, Z_{2,i})'$ where $Z_{1,i} = 1$ is a constant and $Z_{2,i}$ is uniformly distributed on a set \mathcal{Z}_2 of K equally spaced points on $[-5, 5]$. For a true parameter $\theta_0 = (1, 2)'$, we then generate:

$$Y_i = Z_i' \theta_0 + \epsilon_i \quad i = 1, \dots, n, \quad (63)$$

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

$$\begin{aligned} Y_{L,i} &= Y_i - C - V_i Z_i^2 & i = 1, \dots, n \\ Y_{U,i} &= Y_i + C + V_i Z_i^2 & i = 1, \dots, n, \end{aligned} \quad (64)$$

where $C > 0$ and V_i is uniformly distributed on $[0, 0.2]$ independently of (Y_i, Z_i) . As discussed in Example 2.1, $\Theta_0(P_0)$ consists of all $\theta \in \Theta$ such that $E[Y_{L,i}|Z_i] \leq Z_i' \theta \leq E[Y_{U,i}|Z_i]$ almost surely. All our reported simulation results are based on 5000 replications.

Our Monte Carlo experiment is designed to examine the robustness of the estimator to the two free parameters K and C . Since $d_F = 2K$, the constant K determines the number of constraints. Increasing K implies the “flat faces” of $\Theta_0(P_0)$ determined by the linear constraint becomes shorter, which heuristically means $\Theta_0(P_0)$ is closer to violating Assumption 3.4(iii); see Figure 1. Therefore, we interpret designs with large K as close to the analogue of an intersection bounds model in a scalar setting. Similarly, C controls the diameter of the identified set, with point-identification occurring at $C = 0$. Hence, we consider designs with small C as being closer to violating Assumption 3.1(iii).

Since in this context all constraints are linear, the support function has the dual representation:

$$\nu(p, \Theta_0(\hat{P}_n)) = \min_{w \in \mathbf{R}_+^{2d_F}} \{ \langle w, F_L(\frac{1}{n} \sum_{i=1}^n m_L(X_i)) \rangle \text{ s.t. } A'w = p \}, \quad (65)$$

where A and $v \mapsto F_L(v)$ are as defined in Example 2.1. Moreover, the minimizers of (65) are the Lagrange multipliers $\lambda(p, \hat{P}_n)$ of the primal problem that defines $\nu(p, \Theta_0(\hat{P}_n))$.¹⁷ Therefore, by (38) and direct calculation, solving (65) suffices for computing the bootstrap process G_n^* by:

$$G_n^*(p) = -\lambda(p, \hat{P}_n)' \nabla F_L(\frac{1}{n} \sum_{i=1}^n m_L(X_i)) \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i \{ m_L(X_i) - \frac{1}{n} \sum_{i=1}^n m_L(X_i) \}. \quad (66)$$

In our simulations we employ Rademacher weights for this construction – i.e., $P(W_i = 1) = P(W_i = -1) = 1/2$. We emphasize the low computational cost of this exercise, which does not require us to solve for the optimal values $\hat{\theta}(p)$ nor recompute the support function in each bootstrap iteration.

An issue to note, however, is that the dual representation in (65) is only valid for samples in

¹⁷This problem is easily solvable by standard packages. We employ the open software Matlab toolboxes YALMIP and MPT, available at <http://users.isy.liu.se/johan1/yalmip/> and <http://control.ee.ethz.ch/~mpt/>.

which $\Theta_0(\hat{P}_n)$ is nonempty. As mentioned in Remark 4.2, this is likely to occur in specifications where the parameter is near point identification (C is small) and/or the number of constraints ($2K$) is large. In replications for which $\Theta_0(\hat{P}_n)$ is empty, we follow Remark 4.2 setting $\epsilon_n = 1_{d_F} k/n^{\frac{2}{3}}$ where $1_{d_F} \in \mathbf{R}^{d_F}$ is a vector of ones and k is the smallest integer (possibly zero), for which $\tilde{\nu}_n(\cdot, \Theta(\hat{P}_n))$ is well defined. Table 1 reports the proportion of simulation replications for which this modification takes place. As expected, the need to introduce slackness in the constraints is most frequent for the smallest value of C ($C = 0.1$) and largest value of K ($K = 15$) we consider.

Tables 2-4 report the average of the Hausdorff distance and two directed Hausdorff distance measures between $\hat{\Theta}_n$ and $\Theta_0(P_0)$. We call $\vec{d}_H(\hat{\Theta}_n, \Theta_0(P_0))$ and $\vec{d}_H(\Theta_0(P_0), \hat{\Theta}_n)$ the *inner Hausdorff distance* and *outer Hausdorff distance* respectively. As expected, the average Hausdorff distance decreases with sample size across all specifications. Decrements in C yield almost no change in performance, with the exception of specifications where slackness needed to be introduced into the constraints. In contrast, increments in K significantly alter the average Hausdorff measures with asymmetric effects on the inner and outer Hausdorff distances. As K gets large, the inner Hausdorff distance starts decreasing beyond some value of K while the outer Hausdorff distance monotonically increases with K . This is in accord with sample analog estimators having a non-negligible finite sample bias toward the interior of the identified set in intersection bound problems (Manski and Pepper, 2000). This bias renders the estimator more likely to be inside the identified set, lowering the inner Hausdorff distance and inflating the outer Hausdorff distance.

Tables 5 and 6 report the coverage probabilities of the one sided and two sided confidence sets under different values of (n, K, C) and a nominal level of 0.95. Our benchmark specification is $n = 1,000$, $K = 5$, and $C = 1$, for which both confidence sets achieve coverage probabilities that are practically at the nominal level. Coverage probabilities remain very accurate under smaller sample sizes provided $(K, C) = (5, 1)$. These results suggest that our confidence sets have good finite sample properties when violations of Assumptions 3.4(iii) and 3.1(iii) are not of concern.

As expected, increasing K makes Assumption 3.1(iii) more tenuous, and the performance of the confidence intervals deteriorates across specifications for (n, C) . For example, for $n = 1000$, the coverage probability of the one sided confidence set changes from 0.950 to 0.921 when K changes from 5 to 10 and to 0.891 when K changes to 15. This suggests that when many moment inequalities are present, inference methods that do not explicitly take into account this feature may provide a poor approximation; see Menzel (2008). Analogously, decreasing C moves $\Theta_0(P_0)$ closer to violating Assumption 3.4(iii) which affects the coverage probabilities of the constructed confidence regions by making them conservative across most specifications.

7 Conclusion

This paper develops what is, to our knowledge, the first semiparametric efficiency analysis of partially identified models. We obtained conditions under which the support function of the identified

set is a parameter that admits for regular estimators and characterized the semiparametric efficiency bound for estimating it. The intuitive plug-in estimator was shown to be semiparametrically efficient, and the associated set estimator was shown to minimize asymptotic risk for a wide class of loss functions based on Hausdorff distance. We further developed a bootstrap procedure for estimating the asymptotic distribution of the efficient estimator, which enabled us to construct diverse confidence regions previously considered in the literature. It is worth noting, however, that in principle the efficient estimator may be employed in a different manner, for example by constructing confidence regions that studentize the estimated support function.

As in standard maximum likelihood problems, the analysis is local in nature and the applicability of the results in finite samples depends on how close the necessary assumptions are to being violated. Our simulations confirm that, in accord to the theory, the performance of the procedure is excellent under favorable conditions, but deteriorates as we approximate the analogue of an intersection bounds problem. A fruitful venue for future research would therefore be to develop an adaptive hybrid procedure that in a data dependant way selects whether to employ the semiparametric efficient estimator or one more suitable to irregular settings.

APPENDIX A - Notation and Definitions

The following is a list of the notation and definitions that will be used throughout the appendix.

$a \lesssim b$	$a \leq Mb$ for some constant M that is universal in the context of the proof.
$\ \cdot\ _F$	The Frobenius norm $\ A\ _F^2 \equiv \text{trace}\{A'A\}$.
$\ \cdot\ _o$	The operator norm for linear mappings.
\mathbf{M}	The set of Borel probability measures on $\mathcal{X} \subseteq \mathbf{R}^{d_x}$.
$N(Q)$	A subset of \mathbf{M} that contains Q in its interior (relative to the τ -topology).
$N(\epsilon, \mathcal{F}, \ \cdot\)$	Covering numbers of size ϵ for \mathcal{F} under norm $\ \cdot\ $.
$N_{[\]}(\epsilon, \mathcal{F}, \ \cdot\)$	Bracketing numbers of size ϵ for \mathcal{F} under norm $\ \cdot\ $.

APPENDIX B - Proof of Theorem 3.2

Lemma B.1. *Under Assumptions 3.2(i), 3.3(i) and 3.3(iii) there is $N(P_0) \subseteq \mathbf{M}$ such that if $(\theta_0, Q_0) \in \Theta \times N(P_0)$ are such that $F(\int m(x, \theta_0)dQ_0(x)) < 0$, then there is $N(Q_0) \subseteq \mathbf{M}$ with $F(\int m(x, \theta_0)dQ(x)) < 0$ for all $Q \in N(Q_0)$.*

Proof: By Assumption 3.3(i), there is a $N(P_0) \subseteq \mathbf{M}$ such that for all $(\theta, Q) \in \Theta \times N(P_0)$, $\int m(x, \theta)dQ(x)$ is a continuity point of F . Since $m : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^{d_m}$ is bounded by Assumption 3.2(i), the mapping $Q \mapsto \int m(x, \theta_0)dQ(x)$ is continuous with respect to the τ -topology. Therefore, for any $Q \in N(P_0)$, $Q \mapsto F(\int m(x, \theta_0)dQ(x))$ is continuous in the τ -topology and the claim of the Lemma follows from $Q_0 \in N(P_0)$ and $F(\int m(x, \theta_0)dQ_0(x)) < 0$. ■

Corollary B.1. *If Assumptions 3.1(iii), 3.2(i), 3.3(i) and 3.3(iii) hold, then there exists a $\theta_0 \in \Theta$ and a neighborhood $N(P_0) \subseteq \mathbf{M}$ such that for all $Q \in N(P_0)$ we have $F(\int m(x, \theta_0)dQ(x)) < 0$.*

Proof: The claim follows immediately from Lemma B.1 applied to $Q_0 = P_0$ and noting that the existence of a $\theta_0 \in \Theta$ such that $F(\int m(x, \theta_0)dP_0(x)) < 0$ is guaranteed by Assumption 3.1(iii). ■

Lemma B.2. *Let $f : \mathcal{X} \times \Theta \rightarrow \mathbf{R}$ be such that $f(x, \theta)$ is bounded in $(x, \theta) \in \mathcal{X} \times \Theta$ and $\theta \mapsto f(x, \theta)$ is equicontinuous in $x \in \mathcal{X}$. If Assumption 3.1(i) holds and $\{Q_\alpha\}_{\alpha \in A}$ is a net in \mathbf{M} with $Q_\alpha \rightarrow Q_0$, then uniformly in $\theta \in \Theta$:*

$$\int f(x, \theta)dQ_\alpha(x) \rightarrow \int f(x, \theta)dQ_0(x) .$$

Proof: Fix $\epsilon > 0$ and let $N_\delta(\theta) \equiv \{\tilde{\theta} \in \Theta : \|\theta - \tilde{\theta}\| < \delta\}$. By equicontinuity, for every $\theta \in \Theta$ there is a $\delta(\theta)$ with:

$$\sup_{x \in \mathcal{X}, \tilde{\theta} \in N_{\delta(\theta)}(\theta)} |f(x, \theta) - f(x, \tilde{\theta})| < \epsilon . \quad (67)$$

By compactness of Θ , there then exists a finite collection $\{\theta_1, \dots, \theta_K\}$ such that $\{N_{\delta(\theta_i)}(\theta_i)\}_{i=1}^K$ covers Θ . Hence,

$$\left| \int f(x, \theta)dQ_\alpha(x) - \int f(x, \theta)dQ_0(x) \right| \leq 2\epsilon + \max_{1 \leq i \leq K} \left| \int f(x, \theta_i)(dQ_\alpha(x) - dQ_0(x)) \right| \quad (68)$$

for any $\theta \in \Theta$. Since ϵ is arbitrary and $\max_{1 \leq i \leq K} \left| \int f(x, \theta_i)(dQ_\alpha(x) - dQ_0(x)) \right| \rightarrow 0$ due to $f(x, \theta)$ being measurable and bounded for all θ and $Q_\alpha \rightarrow Q_0$ in the τ -topology, the claim of the Lemma then follows from (68). ■

Lemma B.3. *If Assumptions 3.1(i), 3.1(iii), 3.2(i)-(ii), 3.3(i)-(iii) and 3.4(ii) hold, then there is a neighborhood $N(P_0) \subseteq \mathbf{M}$ such that the correspondence $Q \mapsto \Theta_0(Q)$ is well defined and continuous at all $Q \in N(P_0)$.*

Proof: By Assumption 3.4(ii) and Corollary B.1, there exists a $N(P_0) \subseteq \mathbf{M}$ such that for all $Q \in N(P_0)$ and $1 \leq i \leq d_F$ the functions $\theta \mapsto F^{(i)}(\int m(x, \theta)dQ(x))$ are convex at all $\theta \in \Theta$ and $F(\int m(x, \theta_0)dQ(x)) < 0$ for some $\theta_0 \in \Theta$. Thus, in what follows we let $\Theta_0(Q)$ be a convex set with nonempty interior.

We first establish $Q \mapsto \Theta_0(Q)$ is lower hemicontinuous at any $Q_0 \in N(P_0)$. By Theorem 17.19 in Aliprantis and Border (2006), it suffices to show that for any $\theta^* \in \Theta_0(Q_0)$ and net $\{Q_\alpha\}_{\alpha \in A}$ with $Q_\alpha \rightarrow Q_0$, there exists a subnet $\{Q_{\alpha_\beta}\}_{\beta \in B}$ and net $\{\theta_\beta\}_{\beta \in B}$ such that $\theta_\beta \in \Theta_0(Q_{\alpha_\beta})$ for all $\beta \in B$ and $\theta_\beta \rightarrow \theta^*$. If $\theta^* \in \Theta_0^\circ(Q_0)$, then $F(\int m(x, \theta^*) dQ_0(x)) < 0$ and hence by Lemma B.1 and $Q_\alpha \rightarrow Q_0$, there exists α_0 such that $\theta^* \in \Theta_0(Q_\alpha)$ for all $\alpha \geq \alpha_0$. Therefore, defining $B \equiv \{\alpha \in A : \alpha \geq \alpha_0\}$, $Q_{\alpha_\beta} = Q_\beta$ and setting $\theta_\beta = \theta^*$ we obtain $\{Q_{\alpha_\beta}\}_{\beta \in B}$ is a subnet with $\theta_\beta \in Q_{\alpha_\beta}$ and trivially satisfies $\theta_\beta \rightarrow \theta^*$. Suppose on the other hand $\theta^* \in \partial\Theta_0(Q_0)$. Since $\Theta_0(Q_0)$ is convex with nonempty interior, there is a sequence $\tilde{\theta}_k$ with $\tilde{\theta}_k \rightarrow \theta^*$ and $\tilde{\theta}_k \in \Theta_0^\circ(Q_0)$ for all k . By Lemma B.1, there then exists a $\alpha_{0,k}$ such that $\tilde{\theta}_k \in \Theta_0(Q_\alpha)$ for all $\alpha \geq \alpha_{0,k}$. Let $B \equiv A \times \mathbb{N}$ and for any $\beta = (\alpha, k)$ let $\alpha_\beta = \tilde{\alpha}$ for some $\tilde{\alpha} \in A$ with $\tilde{\alpha} \geq \alpha$ and $\tilde{\alpha} \geq \alpha_{0,k}$ and $\theta_\beta = \tilde{\theta}_k$. $\{Q_{\alpha_\beta}\}_{\beta \in B}$ is then a subnet of $\{Q_\alpha\}_{\alpha \in A}$ with $\theta_\beta \in \Theta_0(Q_{\alpha_\beta})$ and $\theta_\beta \rightarrow \theta^*$.

Next, we show that $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous at any $Q_0 \in N(P_0)$. By Theorem 17.16 in Aliprantis and Border (2006), it suffices to show that any net $\{Q_\alpha, \theta_\alpha\}_{\alpha \in A}$ such that $Q_\alpha \rightarrow Q_0$ and $\theta_\alpha \in \Theta_0(Q_\alpha)$ for all $\alpha \in A$ is such that $\{\theta_\alpha\}_{\alpha \in A}$ has a limit point $\theta^* \in \Theta_0(Q_0)$. To this end, first observe that for any $\theta_1, \theta_2 \in \Theta$:

$$\lim_{\theta_1 \rightarrow \theta_2} F\left(\int m(x, \theta_1) dQ_0(x)\right) = F\left(\int m(x, \theta_2) dQ_0(x)\right), \quad (69)$$

due to Assumptions 3.2(i)-(ii), 3.3(i), 3.3(iii) and the dominated convergence theorem. Next note that by compactness of Θ , there exists a subnet $\{\theta_{\alpha_\beta}\}_{\beta \in B}$ such that $\theta_{\alpha_\beta} \rightarrow \theta^*$ for some $\theta^* \in \Theta$. Moreover, since $x \mapsto m(x, \theta)$ is bounded in (x, θ) and equicontinuous in θ by Assumption 3.2(ii), Lemma B.2 and Assumptions 3.3(ii) imply:

$$F\left(\int m(x, \theta) dQ_{\alpha_\beta}(x)\right) \rightarrow F\left(\int m(x, \theta) dQ_0(x)\right) \quad (70)$$

uniformly in $\theta \in \Theta$. Therefore, since $\theta_{\alpha_\beta} \in \Theta_0(Q_{\alpha_\beta})$ for all $\beta \in B$, we obtain from results (69) and (70) that:

$$0 \geq F\left(\int m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x)\right) \rightarrow F\left(\int m(x, \theta^*) dQ_0(x)\right). \quad (71)$$

Thus, $\theta^* \in \Theta_0(Q_0)$ and upper hemicontinuity is established. Since, as argued, $Q \mapsto \Theta_0(Q)$ is also lower hemicontinuous, the claim of the Lemma immediately follows. ■

Corollary B.2. *Under Assumptions 3.1(i)-(iii), 3.2(i)-(ii), 3.3(i), 3.3(iii) and 3.4(ii), there exists a neighborhood $N(P_0) \subseteq \mathbf{M}$ so that $\Theta_0(Q) \subset \Theta^\circ$ for all $Q \in N(P_0)$.*

Proof: Since $\theta \mapsto F(\int m(x, \theta) dP_0(x))$ is continuous in $\theta \in \Theta$, as shown in (69), it follows that $\Theta_0(P_0)$ is closed. Hence, since $\partial\Theta$ is closed as well and $\Theta_0(P_0) \cap \partial\Theta = \emptyset$ by Assumption 3.1(ii), they must be well separated:

$$\inf_{\theta_1 \in \Theta_0(P_0)} \inf_{\theta_2 \in \partial\Theta} \|\theta_1 - \theta_2\| > 0. \quad (72)$$

Therefore, there exists an open set U such that $\Theta_0(P_0) \subset U \subset \Theta^\circ$. Since by Lemma B.3, the correspondence $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous at P_0 , there then exists a $N(P_0) \subseteq \mathbf{M}$ such that for all $Q \in N(P_0)$ we have $\Theta_0(Q) \subset U \subset \Theta^\circ$; see Definition 17.2 in Aliprantis and Border (2006). ■

Lemma B.4. *Let Assumptions 3.1(i), 3.1(iii), 3.2(i)-(ii), 3.3(i), 3.3(iii), 3.4(ii) hold and define the correspondence:*

$$\Xi(p, Q) \equiv \arg \max_{\theta \in \Theta} \{ \langle p, \theta \rangle \text{ s.t. } F\left(\int m(x, \theta) dQ(x)\right) \leq 0 \}. \quad (73)$$

Then there is $N(P_0) \subseteq \mathbf{M}$ with $(p, Q) \mapsto \Xi(p, Q)$ non-empty, compact and upper hemicontinuous on $\mathbb{S}^{d_\theta} \times N(P_0)$.

Proof: By Lemma B.3, there exists a $N(P_0) \subseteq \mathbf{M}$ such that $Q \mapsto \Theta_0(Q)$ is well defined and continuous on $N(P_0)$. Since by (69) the set $\Theta_0(Q) \subseteq \Theta$ is closed, Assumption 3.1(i) implies $\Theta_0(Q)$ is compact. Hence, $\Xi(p, Q)$ is well defined as the maximum is indeed attained for all $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P_0)$. Continuity of $Q \mapsto \Theta_0(Q)$ and Theorem 17.31 in Aliprantis and Border (2006) then imply $(p, Q) \mapsto \Xi(p, Q)$ is compact valued and upper hemicontinuous. ■

Lemma B.5. *Under Assumptions 3.1(i), 3.1(iii), 3.2(i)-(ii), 3.3(i)-(iii) and 3.4(ii)-(iii) there exists $N(P_0) \subseteq \mathbf{M}$ such that for all $Q \in N(P_0)$ we have $\#\mathcal{A}(\theta, Q) \leq d_\theta$ for all $\theta \in \Theta_0(Q)$.*

Proof: We prove the claim by contradiction. Let \mathfrak{N}_{P_0} denote the neighborhood system of P_0 with the direction $V \succeq W$ whenever $V \subseteq W$, which forms a directed set. Let $A = \mathfrak{N}_{P_0}$ and note that if the claim of the Lemma fails to hold, then there exists a net $\{Q_\alpha, \theta_\alpha\}_{\alpha \in A}$ such that $Q_\alpha \rightarrow P_0$, $\theta_\alpha \in \Theta_0(Q_\alpha)$ and $\#\mathcal{A}(\theta_\alpha, Q_\alpha) > d_\theta$ for all $\alpha \in A$. By Lemma B.3, $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous, while $\Theta_0(Q_\alpha)$ is closed by (69) and hence compact by Assumption 3.1(i). Therefore, by Theorem 17.16 in Aliprantis and Border (2006) there is a subnet $\{Q_{\alpha_\beta}, \theta_{\alpha_\beta}\}_{\beta \in B}$ with:

$$\#\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) > d_\theta \quad (\theta_{\alpha_\beta}, Q_{\alpha_\beta}) \rightarrow (\theta^*, P_0) \quad (74)$$

for some $\theta^* \in \Theta_0(P_0)$. Moreover, since for any index $i \in \mathcal{A}^c(\theta^*, P_0)$ results (69) and (70) in turn imply that:

$$F^{(i)}\left(\int m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x)\right) \rightarrow F^{(i)}\left(\int m(x, \theta^*) dP_0(x)\right) < 0, \quad (75)$$

it follows that there is a β_0 such that if $\beta \geq \beta_0$ then constraints that are inactive under (θ^*, P_0) are also inactive under $(\theta_{\alpha_\beta}, Q_{\alpha_\beta})$. We conclude that for $\beta \geq \beta_0$, $\mathcal{A}^c(\theta^*, P_0) \subseteq \mathcal{A}^c(\theta_{\alpha_\beta}, Q_{\alpha_\beta})$ and thus we obtain that:

$$\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) \subseteq \mathcal{A}(\theta^*, P_0). \quad (76)$$

Since by Assumption 3.4(iii) $\#\mathcal{A}(\theta^*, P_0) \leq d_\theta$, result (76) contradicts (74) and the claim of the Lemma follows. ■

Lemma B.6. *Under Assumptions 3.1(i), 3.1(iii), 3.2, 3.3, 3.4(ii) and 3.4(iv) there is a $N(P_0) \subseteq \mathbf{M}$ so that $\{\nabla F^{(i)}(\int m(x, \theta) dQ(x)) \int \nabla_\theta m(x, \theta) dQ(x)\}_{i \in \mathcal{A}(\theta, Q)}$ are linearly independent for all $\theta \in \Theta_0(Q)$ and $Q \in N(P_0)$.*

Proof: The proof is by contradiction. As in the proof of Lemma B.5, let \mathfrak{N}_{P_0} be the neighborhood system of P_0 with direction $V \succeq W$ whenever $V \subseteq W$. If the Lemma fails to hold, then for $A = \mathfrak{N}_{P_0}$ there exists a net $\{Q_\alpha, \theta_\alpha\}_{\alpha \in A}$ such that $Q_\alpha \rightarrow P_0$, $\theta_\alpha \in \Theta_0(Q_\alpha)$ and the vectors $\{\nabla F^{(i)}(\int m(x, \theta_\alpha) dQ_\alpha(x)) \int \nabla_\theta m(x, \theta_\alpha) dQ_\alpha(x)\}_{i \in \mathcal{A}(\theta_\alpha, Q_\alpha)}$ are not linearly independent for all $\alpha \in A$. Since by Lemma B.3 the correspondence $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous, we may pass to a subnet $\{Q_{\alpha_\beta}, \theta_{\alpha_\beta}\}_{\beta \in B}$ such that $(Q_{\alpha_\beta}, \theta_{\alpha_\beta}) \rightarrow (P_0, \theta^*)$ with $\theta^* \in \Theta_0(P_0)$. Moreover, since as shown in (76), there is a β_0 such that $\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) \subseteq \mathcal{A}(\theta^*, P_0)$ for $\beta \geq \beta_0$, in establishing a contradiction it suffices to show $\{\nabla F^{(i)}(\int m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x)) \int \nabla_\theta m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x)\}_{i \in \mathcal{A}(\theta^*, P_0)}$ are linearly independent for some $\beta \geq \beta_0$.

Towards this end, notice that Assumptions 3.2(ii)-(iii) and Lemma B.2 imply that uniformly in $\theta \in \Theta$:

$$\int \nabla_\theta m(x, \theta) dQ_{\alpha_\beta}(x) \rightarrow \int \nabla_\theta m(x, \theta) dQ_0(x). \quad (77)$$

Moreover, since $\nabla_\theta m$ is uniformly bounded and continuous in θ , the dominated convergence theorem and (77) imply:

$$\int \nabla_\theta m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x) \rightarrow \int \nabla_\theta m(x, \theta^*) dQ_0(x). \quad (78)$$

Similarly, since $v \mapsto \nabla F(v)$ is uniformly continuous on $v \in V$ by Assumption 3.3(ii) and $\int m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x) \in V$ by Assumption 3.3(iii), we obtain from (78) and applying Lemma B.2 to $\theta \mapsto m(x, \theta)$ that:

$$\nabla F\left(\int m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x)\right) \int \nabla_\theta m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x) \rightarrow \nabla F\left(\int m(x, \theta^*) dP_0(x)\right) \int \nabla_\theta m(x, \theta^*) dP_0(x). \quad (79)$$

However, by Assumption 3.4(iv), the vectors $\{\nabla F^{(i)}(\int m(x, \theta^*) dP_0(x)) \int \nabla_\theta m(x, \theta^*) dP_0(x)\}_{i \in \mathcal{A}(\theta^*, P_0)}$ are linearly independent and hence by (79), so must $\{\nabla F^{(i)}(\int m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x)) \int \nabla_\theta m(x, \theta_{\alpha_\beta}) dQ_{\alpha_\beta}(x)\}_{i \in \mathcal{A}(\theta^*, P_0)}$ for $\beta \geq \beta_1$ and some $\beta_1 \in B$. Thus, the contradiction is established and the claim of the Lemma follows. ■

Lemma B.7. *If Assumptions 3.1(i)-(iii), 3.2(i)-(iii), 3.3(i)-(iii) and 3.4(ii)-(iv) hold, then there exists a $N(P_0) \subseteq \mathbf{M}$ such that for all $Q \in N(P_0)$ and $p \in \mathbb{S}^{d_\theta}$ there is a unique $\lambda(p, Q) \in \mathbf{R}^{d_F}$ satisfying the following:*

$$\sup_{\theta \in \Theta_0(Q)} \langle p, \theta \rangle = \sup_{\theta \in \Theta} \langle p, \theta \rangle + \lambda(p, Q)' F\left(\int m(x, \theta) dQ(x)\right). \quad (80)$$

Proof: By Corollary B.1 and Assumption 3.4(ii) there is a $N(P_0) \subseteq \mathbf{M}$ such that for all $Q \in N(P_0)$ there is a $\theta_0 \in \Theta$ with $F(\int m(x, \theta_0) dQ(x)) < 0$ and in addition $\theta \mapsto F^{(i)}(\int m(x, \theta) dQ(x))$ is convex in $\theta \in \Theta$ for all $1 \leq i \leq d_F$.

Moreover, since by compactness of Θ the supremum on the left hand side of (80) is finite, it follows that the problem:

$$\sup_{\theta \in \Theta} \langle p, \theta \rangle \quad \text{s.t.} \quad F\left(\int m(x, \theta) dQ(x)\right) \leq 0 \quad (81)$$

satisfies the conditions of Corollary 28.2.1 in Rockafellar (1970) for all $Q \in N(P_0)$ and all $p \in \mathbb{S}^{d_\theta}$. We can therefore conclude that the equality in (80) holds for some $\lambda(p, Q) \in \mathbf{R}^{d_F}$.

In order to show that $\lambda(p, Q)$ is unique for all $p \in \mathbb{S}^{d_\theta}$ and $Q \in N(P_0)$ for some $N(P_0) \subseteq \mathbf{M}$, note that by Lemma B.4 the supremum in the convex problem in (81) is attained. In turn, for $\Xi(p, Q)$ as defined in (73), Theorem 8.3.1 in Luenberger (1969) implies any $\theta^* \in \Xi(p, Q)$ is also a maximizer of the dual problem. Since by Corollary B.2, $\Xi(p, Q) \subset \Theta^\circ$ for all $p \in \mathbb{S}^{d_\theta}$ and Q in a neighborhood of P_0 , it follows that for any $\theta^* \in \Xi(p, Q)$:

$$p' + \lambda(p, Q)' \nabla F\left(\int m(x, \theta^*) dQ(x)\right) \int \nabla_\theta m(x, \theta^*) dQ(x) = 0, \quad (82)$$

where the exchange of order of integration and differentiation is warranted by the mean value theorem, the dominated convergence theorem and Assumption 3.2(ii). Result (82) represents a linear equation in $\lambda(p, Q) \in \mathbf{R}^{d_F}$. However, by the complementary slackness conditions $\lambda^{(i)}(p, Q) = 0$, for any $i \in \mathcal{A}^c(\theta^*, Q)$. Therefore, the linear system in equation (82) can be reduced to d_θ equations and $\#\mathcal{A}(\theta^*, Q)$ unknowns. The uniqueness of the solution then follows from the linear independence of the vectors $\{\nabla F^{(i)}(\int m(x, \theta^*) dQ(x)) \int \nabla_\theta m(x, \theta^*) dQ(x)\}_{i \in \mathcal{A}(\theta^*, Q)}$ by Lemma B.6 and that $\#\mathcal{A}(\theta^*, Q) \leq d_\theta$ by Lemma B.5. We conclude $\lambda(p, Q)$ is uniquely defined and the Lemma follows. ■

Lemma B.8. *Let Assumptions 3.1(i), 3.1(iii), 3.2(i)-(ii), 3.3(i), 3.3(iii), 3.4(ii) hold and $\Xi(p, Q)$ be as in (73). Then there is a neighborhood $N(P_0) \subseteq \mathbf{M}$ such that for all $Q \in N(P_0)$ and $p \in \mathbb{S}^{d_\theta}$ for which $\Xi(p, Q)$ is not a singleton, $F^{(i)}(\int m_S(x, \theta_L) dQ(x)) < 0$ for all $1 \leq i \leq d_{F_S}$ for some $\theta_L \in \Xi(p, Q)$.*

Proof: Note that by Assumption 3.4(ii) and Lemma B.4 there exists a $N(P_0) \subseteq \mathbf{M}$ such that for all $Q \in N(P_0)$ and $1 \leq i \leq d_{F_S}$, the functions $\theta \mapsto F^{(i)}(\int m_S(x, \theta) dQ(x))$ are strictly convex and $\Xi(p, Q)$ is well defined for all $p \in \mathbb{S}^{d_\theta}$. Suppose $\theta_1, \theta_2 \in \Xi(p, Q)$ with $\theta_1 \neq \theta_2$. By defining $\theta_L = \lambda\theta_1 + (1-\lambda)\theta_2$ with $\lambda \in (0, 1)$ we obtain for any $1 \leq i \leq d_{F_S}$:

$$F^{(i)}\left(\int m_S(x, \theta_L) dQ(x)\right) < \lambda F^{(i)}\left(\int m_S(x, \theta_1) dQ(x)\right) + (1-\lambda) F^{(i)}\left(\int m_S(x, \theta_2) dQ(x)\right) \leq 0 \quad (83)$$

where the first inequality is a result of the strict convexity and the second inequality follows from $\theta_1, \theta_2 \in \Theta_0(Q)$. However, by Assumption 3.4(ii) $\Theta_0(Q)$ is convex and hence $\theta_L \in \Theta_0(Q)$. Since $\langle p, \theta_L \rangle = \lambda \langle p, \theta_1 \rangle + (1-\lambda) \langle p, \theta_2 \rangle$, we must have $\theta_L \in \Xi(p, Q)$ and therefore (83) establishes the Lemma. ■

Lemma B.9. *Let Assumptions 3.1(i)-(iii), 3.2(i)-(iii), 3.3(i)-(iii) and 3.4(ii)-(iv) hold and $\lambda(p, Q)$ be as in (80). Then, there exists a $N(P_0) \subseteq \mathbf{M}$ such that $\|\lambda(p, Q)\|$ is uniformly bounded in $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P_0)$.*

Proof: We establish the claim by contradiction. Let \mathfrak{N}_{P_0} denote the neighborhood system of P_0 with direction $V \succeq W$ whenever $V \subseteq W$ and \mathbb{N} be the natural numbers and note $\mathfrak{N}_{P_0} \times \mathbb{N}$ then forms a directed set. Setting $A = \mathfrak{N}_{P_0} \times \mathbb{N}$ and letting $\alpha = (V(\alpha), k(\alpha)) \in A$ we may then find a net $\{Q_\alpha, p_\alpha, \theta_\alpha\}_{\alpha \in A}$ such that for all $\alpha \in A$ we obtain:

$$\|\lambda(p_\alpha, Q_\alpha)\| > k(\alpha) \quad Q_\alpha \rightarrow P_0 \quad p_\alpha \in \mathbb{S}^{d_\theta} \quad \theta_\alpha \in \Xi(p_\alpha, Q_\alpha), \quad (84)$$

for all $\alpha \in A$ and $\Xi(p, Q)$ as in (73). However, by (i) $(p, Q) \mapsto \Xi(p, Q)$ being upper hemicontinuous and compact valued and (ii) \mathbb{S}^{d_θ} being compact, we may pass to a subnet $\{Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}\}_{\beta \in B}$ such that:

$$(Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}) \rightarrow (P_0, p^*, \theta^*) \quad \text{for some } (p^*, \theta^*) \in \mathbb{S}^{d_\theta} \times \Xi(p^*, P_0). \quad (85)$$

Moreover, since the number of constraints is finite, there is a set of indices $\mathcal{C} \subseteq \{1, \dots, d_F\}$ such that for every $\beta_0 \in B$ there exists a $\beta \geq \beta_0$ with $\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) = \mathcal{C}$. Letting $\Gamma \equiv B$ we may then set $\alpha_{\beta_\gamma} = \alpha_{\tilde{\beta}}$ for some $\tilde{\beta} \geq \beta$ and satisfying $\mathcal{A}(\theta_{\alpha_{\tilde{\beta}}}, Q_{\alpha_{\tilde{\beta}}}) = \mathcal{C}$. In this way, we obtain a subnet which, for simplicity, we denote $\{Q_{\alpha_\gamma}, p_{\alpha_\gamma}, \theta_{\alpha_\gamma}\}_{\gamma \in \Gamma}$ with:

$$\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\| > k(\alpha_\gamma) \quad (Q_{\alpha_\gamma}, p_{\alpha_\gamma}, \theta_{\alpha_\gamma}) \rightarrow (P_0, p^*, \theta^*) \quad \mathcal{A}(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) = \mathcal{C} \quad \forall \gamma \in \Gamma. \quad (86)$$

Next, let $\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma})$ and $\nabla^{\mathcal{C}}F(\int m(x, \theta_{\alpha_\gamma})dQ_{\alpha_\gamma}(x))$ respectively be the $\#\mathcal{C} \times 1$ vector and $\#\mathcal{C} \times d_m$ matrix that stacks components of $\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma}(x))$ and $\nabla F(\int m(x, \theta_{\alpha_\gamma})dQ_{\alpha_\gamma})$ whose indexes belong to \mathcal{C} . Similarly, we also define:

$$M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) \equiv \nabla^{\mathcal{C}}F(\int m(x, \theta_{\alpha_\gamma})dQ_{\alpha_\gamma}(x)) \int \nabla_{\theta}m(x, \theta_{\alpha_\gamma})dQ_{\alpha_\gamma}(x). \quad (87)$$

By Lemmas B.5 and B.6 there is a γ_0 such that $M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})'$ is invertible for all $\gamma \geq \gamma_0$. Therefore, since by the complementary slackness conditions $\lambda^{(i)}(p_{\alpha_\gamma}, Q_{\alpha_\gamma}) = 0$ for all $i \notin \mathcal{C}$, we obtain from result (82) that:

$$\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma}) = -(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})')^{-1}M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})p_{\alpha_\gamma}. \quad (88)$$

Additionally note that since $(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) \rightarrow (\theta^*, P_0)$ as in (85), we obtain from result (79) and definition (87) that:

$$M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})' \rightarrow M(\theta^*, P_0)M(\theta^*, P_0)'. \quad (89)$$

For a symmetric matrix H , let $\xi(H)$ denote its smallest eigenvalue and note $\xi(M(\theta^*, P_0)M(\theta^*, P_0)') > 2\epsilon$ for some $\epsilon > 0$ by Assumptions 3.4(iii)-(iv). Since eigenvalues are continuous under $\|\cdot\|_F$ by Corollary III.2.6 in Bhatia (1997),

$$\xi(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})') > \epsilon, \quad (90)$$

for all $\gamma \geq \gamma_0$ and some $\gamma_0 \in \Gamma$. Since $\lambda^{(i)}(p_{\alpha_\gamma}, Q_{\alpha_\gamma}) = 0$ for all $i \notin \mathcal{C}$, $\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\| = \|\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\|$ and hence:

$$\begin{aligned} \|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\| &= \|\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\| \\ &\leq \{\|(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})')^{-1}\|_o \times \|M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})\|_F \times \|p\|\} \\ &\leq \xi^{-1}(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})') \times \sup_{v \in V} \|\nabla F(v)\|_F \times \sup_{(x, \theta) \in \mathcal{X} \times \Theta} \|\nabla_{\theta}m(x, \theta)\|_F. \end{aligned} \quad (91)$$

However, (90), (91) and Assumptions 3.2(ii), 3.3(ii)-(iii) imply there is a $\gamma_1 \in \Gamma$ such that $\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\|$ is uniformly bounded for all $\gamma \geq \gamma_1$, contradicting (86). The claim of the Lemma therefore follows. ■

Lemma B.10. *Let Assumptions 3.1(i)-(iii), 3.2(i)-(iii), 3.3(i)-(iii) and 3.4(ii)-(iv) hold and $\lambda(p, Q)$ be as in (80). Then, there exists a $N(P_0) \subseteq \mathbf{M}$ such that the function $(p, Q) \mapsto \lambda(p, Q)$ is continuous on $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P_0)$.*

Proof: By Lemmas B.7 and B.9 there exists a $N(P_0) \subseteq \mathbf{M}$ such that $\lambda(p, Q)$ is well defined, unique and uniformly bounded for all $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P_0)$. Therefore, letting $\Lambda \equiv \text{cl}\{\lambda(p, Q) : (p, Q) \in \mathbb{S}^{d_\theta} \times N(P_0)\}$ it follows that Λ is compact in \mathbf{R}^{d_F} . By Lemma B.7 and Theorem 8.6.1 in Luenberger (1969) we then have:

$$\lambda(p, Q) = \arg \min_{\lambda \geq 0} V(\lambda, p, Q) = \arg \min_{\lambda \in \Lambda} V(\lambda, p, Q) \quad V(\lambda, p, Q) \equiv \max_{\theta \in \Theta} \{ \langle p, \theta \rangle + \lambda' F(\int m(x, \theta)dQ(x)) \}. \quad (92)$$

In turn, $(\theta, Q) \mapsto F(\int m(x, \theta)dQ(x))$ is continuous by (69) and (70) and hence compactness of Θ and Theorem 17.31 in Aliprantis and Border (2006) imply $(\lambda, p, Q) \mapsto V(\lambda, p, Q)$ is continuous. Therefore, by (92), compactness of Λ and a second application of Theorem 17.31 in Aliprantis and Border (2006), it follows that $(p, Q) \mapsto \lambda(p, Q)$ is upper hemicontinuous. However, since $(p, Q) \mapsto \lambda(p, Q)$ is a singleton valued correspondence on $\mathbb{S}^{d_\theta} \times N(P_0)$ by Lemma B.7, we conclude that it is in fact a continuous function. ■

Lemma B.11. *Let Assumptions 3.1(i)-(iii), 3.2(i)-(iii), 3.3(i)-(iii) and 3.4(ii)-(iv) hold and $\Xi(p, P_0)$ be as in (73). Then, there exists a Borel measurable selector $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ with $\theta^*(p) \in \Xi(p, P_0)$ for all $p \in \mathbb{S}^{d_\theta}$.*

Proof: By Lemma B.4, $p \mapsto \Xi(p, P_0)$ is upper hemicontinuous in $p \in \mathbb{S}^{d_\theta}$ and hence weakly measurable; see Definition 18.1 in Aliprantis and Border (2006). Since $p \mapsto \Xi(p, P_0)$ is nonempty and compact valued by Lemma B.4, Theorem 18.13 in Aliprantis and Border (2006), implies there is a measurable selector $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ and the Lemma follows. ■

Lemma B.12. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 such that $\int h_\eta^2(x)d\mu(x) = 1$ and $h_0 = \sqrt{dP_0/d\mu}$. If Assumptions 3.1(i)-(iii), 3.2(i)-(iii), 3.3(i)-(iii), 3.4(i)-(iv) hold and P_η satisfies $h_\eta = \sqrt{dP_\eta/d\mu}$, then there exists a neighborhood*

$N \subseteq \mathbf{R}$ of θ such that for all $\eta_0 \in N$, $p \in \mathbb{S}^{d_\theta}$, $\Xi(p, P_\eta)$ as in (73) and $\lambda(p, P_\eta) \in \mathbf{R}^{d_F}$ as in (80), it follows that:

$$\frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta)) \Big|_{\eta=\eta_0} = 2\lambda(p, P_{\eta_0})' \nabla F \left(\int m(x, \theta^*) h_{\eta_0}^2 d\mu(x) \right) \int m(x, \theta^*) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \quad \text{for any } \theta^* \in \Xi(p, P_{\eta_0}).$$

Proof: For any $1 \leq i \leq d_m$ and $\theta \in \Theta$, first observe that by rearranging terms it follows that for any η_0 :

$$\begin{aligned} & \left| \int m^{(i)}(x, \theta) \{h_{\eta_0}^2(x) - h_\eta^2(x) - 2(\eta_0 - \eta)h_{\eta_0}(x)\dot{h}_{\eta_0}(x)\} d\mu(x) \right| \\ &= \left| \int m^{(i)}(x, \theta) \{(h_\eta(x) - h_{\eta_0}(x))^2 + 2h_{\eta_0}(x)(h_\eta(x) - h_{\eta_0}(x) + (\eta_0 - \eta)\dot{h}_{\eta_0}(x))\} d\mu(x) \right| = o(|\eta - \eta_0|) \end{aligned} \quad (93)$$

where the final result holds by $m(x, \theta)$ bounded by Assumption 3.2(i), Cauchy Schwarz, $\|h_\eta - h_{\eta_0}\|_{L_\mu^2}^2 = O(|\eta - \eta_0|^2)$ and $\|h_\eta - h_{\eta_0} - (\eta - \eta_0)\dot{h}_{\eta_0}\|_{L_\mu^2} = o(|\eta - \eta_0|)$ due to $\eta \mapsto h_\eta$ being Fréchet differentiable with derivative \dot{h}_{η_0} at $\eta = \eta_0$. Also note, that $\|h_\eta - h_{\eta_0}\|_{L_\mu^2} = o(1)$ implies $P_\eta \rightarrow P_{\eta_0}$ with respect to the total variation metric, and hence also with respect to the τ -topology. Thus, for η_0 in a neighborhood of zero, result (93) and Assumptions 3.3(i)-(iii) yield:

$$\frac{\partial}{\partial \eta} F \left(\int m(x, \theta) h_\eta^2 d\mu(x) \right) \Big|_{\eta=\eta_0} = 2\nabla F \left(\int m(x, \theta) h_{\eta_0}^2 d\mu(x) \right) \int m(x, \theta) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x). \quad (94)$$

Moreover, since $\eta \mapsto h_\eta$ is continuously Fréchet differentiable, result (70) implies the derivative in (94) is continuous in η_0 . Therefore, Lemma B.7 and Corollary 5 in Milgrom and Segal (2002) then imply that for η in a neighborhood of zero, $\eta \mapsto \nu(p, \Theta_0(P_\eta))$ is directionally differentiable at $\eta = \eta_0$ with derivatives given by:

$$\frac{\partial}{\partial \eta_+} \nu(p, \Theta_0(P_\eta)) \Big|_{\eta=\eta_0} = \max_{\theta^* \in \Xi(p, P_{\eta_0})} 2\lambda(p, P_{\eta_0})' \nabla F \left(\int m(x, \theta^*) h_{\eta_0}^2 d\mu(x) \right) \int m(x, \theta^*) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \quad (95)$$

$$\frac{\partial}{\partial \eta_-} \nu(p, \Theta_0(P_\eta)) \Big|_{\eta=\eta_0} = \min_{\theta^* \in \Xi(p, P_{\eta_0})} 2\lambda(p, P_{\eta_0})' \nabla F \left(\int m(x, \theta^*) h_{\eta_0}^2 d\mu(x) \right) \int m(x, \theta^*) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \quad (96)$$

where $\frac{\partial}{\partial \eta_+}$ and $\frac{\partial}{\partial \eta_-}$ denote right and left derivatives respectively. Therefore, if $p \in \mathbb{S}^{d_\theta}$ is such that $\Xi(p, P_{\eta_0})$ is a singleton, then both (95) and (96) are equal and the claim of the Lemma follows.

Suppose on the other hand that $p \in \mathbb{S}^{d_\theta}$ is such that $\Xi(p, P_{\eta_0})$ is not a singleton. By Lemma B.8 there exists a $\theta_L \in \Xi(p, P_{\eta_0})$ with $F^{(i)}(\int m_S(x, \theta_L) dP_{\eta_0}(x)) < 0$ for all $1 \leq i \leq d_{F_S}$ and hence by the complementary slackness conditions elements of $\lambda^{(i)}(p, P_{\eta_0}) = 0$ for all $1 \leq i \leq d_{F_S}$. Therefore, for any $\theta^* \in \Xi(p, P_{\eta_0})$ we obtain:

$$\begin{aligned} & 2\lambda(p, P_{\eta_0})' \nabla F \left(\int m(x, \theta^*) h_{\eta_0}^2 d\mu(x) \right) \int m(x, \theta^*) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \\ &= -2\lambda_L(p, P_{\eta_0})' \nabla F_L \left(\int m_L(x) h_{\eta_0}^2 d\mu(x) \right) \int m_L(x) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x), \end{aligned} \quad (97)$$

where $\lambda_L(p, P_{\eta_0})$ is the subvector of elements of $\lambda(p, P_{\eta_0})$ corresponding to the linear constraints. Hence, we conclude that the equality of (95) and (96) still holds and the Lemma is established. ■

Lemma B.13. *Let $\eta \mapsto h_\eta$ be a curve in L_μ^2 with $\int h_\eta^2 d\mu(x) = 1$ and $h_0 = \sqrt{dP_0/d\mu}$. If Assumptions 3.1-3.4 hold and P_η satisfies $h_\eta = \sqrt{dP_\eta/d\mu}$, then (i) There is a neighborhood $N \subseteq \mathbf{R}$ of θ such that $\frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta))$ is bounded in $(p, \eta) \in \mathbb{S}^{d_\theta} \times N$, and (ii) $(p, \eta_0) \mapsto \frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta)) \Big|_{\eta=\eta_0}$ is continuous at all (p, η_0) with $p \in \mathbb{S}^{d_\theta}$ and $\eta_0 = 0$.*

Proof: To establish the first claim of the Lemma, notice that by Lemma B.12 and the Cauchy-Schwarz inequality:

$$\left| \frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta)) \Big|_{\eta=\eta_0} \right| \leq 2\|\lambda(p, P_{\eta_0})\| \times \sup_{v \in V} \|\nabla F(v)\|_F \times \sqrt{d_m} \sup_{(x, \theta) \in \mathcal{X} \times \Theta} \|m(x, \theta)\| \times \|\dot{h}_{\eta_0}\|_{L_\mu^2} \times \|h_{\eta_0}\|_{L_\mu^2}, \quad (98)$$

for η_0 in a neighborhood of zero. Since $\|\dot{h}_\eta\|_{L_\mu^2}$ is continuous in η due to $\eta \mapsto h_\eta$ being continuously Fréchet differentiable, it attains a finite maximum in a neighborhood of zero. Thus, $\|\dot{h}_{\eta_0}\|_{L_\mu^2}$ is uniformly bounded and since $\|h_\eta\|_{L_\mu^2} = 1$ for all η , Lemma B.9, Assumptions 3.2(i), 3.3(ii)-(iii) and (98) establish the first claim of the Lemma.

To establish the second claim, let $(p_n, \eta_n) \rightarrow (p_0, \eta_0)$ and select $\theta_n^* \in \Xi(p_n, P_{\eta_n})$ and $\theta^* \in \Xi(p_0, P_{\eta_0})$ for $\Xi(p, P)$

as in (73). Since $\|m(x, \theta)\|$ is uniformly bounded by Assumption 3.2(i), we obtain for any $1 \leq i \leq d_m$:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left| \int m^{(i)}(x, \theta) \{ \dot{h}_{\eta_n}(x) h_{\eta_n}(x) - \dot{h}_{\eta_0}(x) h_{\eta_0}(x) \} d\mu(x) \right| \\ & \leq \sup_{(x, \theta) \in \mathcal{X} \times \Theta} \|m(x, \theta)\| \times \lim_{n \rightarrow \infty} \{ \|\dot{h}_{\eta_n} - \dot{h}_{\eta_0}\|_{L_\mu^2} \|h_{\eta_n}\|_{L_\mu^2} + \|h_{\eta_n} - h_{\eta_0}\|_{L_\mu^2} \|\dot{h}_{\eta_0}\|_{L_\mu^2} \} = 0, \end{aligned} \quad (99)$$

due to the Cauchy-Schwarz inequality, $\eta \mapsto h_\eta$ being continuously Fréchet differentiable and $\|h_\eta\|_{L_\mu^2} = 1$ for all η . If $p_0 \in \mathbb{S}^{d_\theta}$ is such that $\Xi(p_0, P_{\eta_0})$ is a singleton, then the upper hemicontinuity of $\Xi(p, P)$, established in Lemma B.4, implies $\theta_n^* \rightarrow \theta^*$. Therefore, by results (79) and (99) and the dominated convergence theorem:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \nabla F \left(\int m(x, \theta_n^*) h_{\eta_n}^2(x) d\mu(x) \right) \int m(x, \theta_n^*) \dot{h}_{\eta_n}(x) h_{\eta_n}(x) d\mu(x) \\ & = \nabla F \left(\int m(x, \theta^*) h_{\eta_0}^2(x) d\mu(x) \right) \int m(x, \theta^*) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x). \end{aligned} \quad (100)$$

The continuity of $(p, \eta_0) \mapsto \frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta))|_{\eta=\eta_0}$ at values $(p_0, \eta_0) \in \mathbb{S}^{d_\theta} \times \{0\}$ for which $\Xi(p_0, P_{\eta_0})$ is a singleton is then implied by result (100), and Lemmas B.10 and B.12.

To conclude we consider the case when $\Xi(p_0, P_{\eta_0})$ is not a singleton. Let $\lambda(p, P) = (\lambda_S(p, P), \lambda_L(p, P))$ with $\lambda_S(p, P)$ and $\lambda_L(p, P)$ the Lagrange multipliers corresponding to the strictly convex and linear constraints respectively. By Lemma B.8, there exists a $\theta_L \in \Xi(p_0, P_{\eta_0})$ for which $F^{(i)}(\int m_S(x, \theta_L) dP_{\eta_0}(x)) < 0$ for all $1 \leq i \leq d_{F_S}$ and hence the complementary slackness conditions and $\lambda_S(p_n, P_{\eta_n}) \rightarrow \lambda_S(p_0, P_{\eta_0})$ by Lemma B.10 imply that:

$$\lambda_S(p_0, P_{\eta_0}) = 0 \quad \lim_{n \rightarrow \infty} \lambda_S(p_n, P_{\eta_n}) = 0. \quad (101)$$

Hence, by Lemma B.12, (99) and $\|m(x, \theta)\|, \|\nabla F(v)\|_F$ being uniformly bounded by Assumptions 3.2(i), 3.3(ii)-(iii):

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\partial}{\partial \eta} \nu(p_n, \Theta_0(P_{\eta_n})) \Big|_{\eta=\eta_n} \\ & = \lim_{n \rightarrow \infty} -2\lambda_L(p_n, P_{\eta_n})' \nabla F_L \left(\int m_L(x) h_{\eta_n}^2(x) d\mu(x) \right) \int m_L(x) h_{\eta_n}(x) \dot{h}_{\eta_n}(x) d\mu(x) = \frac{\partial}{\partial \eta} \nu(p_0, \Theta_0(P_{\eta_0})) \Big|_{\eta=\eta_0}, \end{aligned} \quad (102)$$

where the final equality follow from Lemma B.10, and results (79) and (99). Hence the function $(p_0, \eta_0) \mapsto \frac{\partial}{\partial \eta} \nu(p_0, \Theta_0(P_\eta))|_{\eta=\eta_0}$ is continuous at points (p_0, η_0) for which $\Xi(p_0, P_{\eta_0})$ is not a singleton as well. ■

Theorem B.1. *Let $\mathbf{P}_\mu \subseteq \mathbf{M}_\mu$ denote the set of $P \in \mathbf{M}$ such that P satisfies Assumptions 3.1(ii)-(iii), 3.3(iii) and 3.4(i)-(iv) with P in place of P_0 and $\mathbf{P} \equiv \{h \in L_\mu^2 : h = \sqrt{dP/d\mu}$ for some $P \in \mathbf{P}_\mu\}$. Under Assumptions 3.1-3.4, the tangent set of \mathbf{P} at $p_0 \equiv \sqrt{dP_0/d\mu}$ is then given by $\dot{\mathbf{P}}_0 = \{h \in L_\mu^2 : \int h(x) p_0(x) d\mu(x) = 0\}$.*

Proof: The proof exploits a construction in Example 3.2.1 of Bickel et al. (1993). First, define the set:

$$\mathbf{T} \equiv \{h \in L_\mu^2 : \int h(x) p_0(x) d\mu(x) = 0\}, \quad (103)$$

and note that by Proposition 3.2.3 in Bickel et al. (1993) we have $\dot{\mathbf{P}}_0 \subseteq \mathbf{T}$. For the reverse inclusion, pick $h \in \mathbf{T}$ and let $\Psi : \mathbf{R} \rightarrow (0, \infty)$ be continuously differentiable, with $\Psi(0) = \Psi'(0) = 1$ and Ψ, Ψ' and Ψ'/Ψ bounded. For $p_0 \equiv \sqrt{dP_0/d\mu}$, define a parametric family of distributions to be pointwise given by:

$$h_\eta^2(x) \equiv b(\eta) p_0^2(x) \Psi \left(\frac{2\eta h(x)}{p_0(x)} \right) \quad b(\eta) \equiv \left[\int \Psi \left(\frac{2\eta h(x)}{p_0(x)} \right) dP_0(x) \right]^{-1}. \quad (104)$$

Employing Proposition 2.1.1 in Bickel et al. (1993) it is straightforward to verify $\eta \mapsto h_\eta$ is a curve on L_μ^2 such that $h_0 = p_0$. Let P_η satisfy $h_\eta = \sqrt{dP_\eta/d\mu}$ and notice $2^{-\frac{1}{2}} \|h_\eta - p_0\|_{L_\mu^2}$ equals the Hellinger distance between P_η and P_0 . Since convergence with respect to the Hellinger distance implies convergence with respect to the τ -topology, it follows by Lemmas B.5, B.6 and Corollaries B.1 and B.2 there exists a neighborhood $N \subseteq \mathbf{R}$ of 0 such that for all $\eta \in N$ Assumptions 3.1(ii)-(iii), 3.3(iii) and 3.4(i)-(iv) hold with P_η in place of P_0 . We conclude $\eta \mapsto h_\eta$ is a regular

parametric submodel. Moreover, by direct calculation, the Fréchet derivative of h_η at $\eta = 0$ satisfies:

$$\dot{h}_0(x) = \frac{1}{2} \frac{b(0)p_0^2(x)\Psi'(0)2h(x)}{p_0(x)p_0(x)} + \frac{1}{2} \frac{b'(0)p_0^2(x)\Psi(0)}{p_0(x)} = h(x), \quad (105)$$

where we have exploited that by the dominated convergence theorem $b'(0) = 2 \int \Psi'(0)h(x)p_0(x)d\mu(x) = 0$ due to $h \in \mathbf{T}$. Hence, from (105) we conclude that $h \in \dot{\mathbf{P}}_0$ and therefore that $\mathbf{T} = \dot{\mathbf{P}}_0$, which establishes the Theorem. ■

Theorem B.2. *Let $\mathbf{P}_\mu \subseteq \mathbf{M}_\mu$ denote the set of $P \in \mathbf{M}$ such that P satisfies Assumptions 3.1(ii)-(iii), 3.3(iii) and 3.4(i)-(iv) with P in place of P_0 and $\mathbf{P} \equiv \{h \in L_\mu^2 : h = \sqrt{dP/d\mu}$ for some $P \in \mathbf{P}_\mu\}$. If Assumptions 3.1-3.4 hold, then the mapping $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ pointwise defined by $\rho(h_\eta) = \nu(\cdot, \Theta_0(P_\eta))$ for $h_\eta = \sqrt{dP_\eta/d\mu}$ is then weakly pathwise differentiable at $p_0 \equiv \sqrt{dP_0/d\mu}$. Moreover, for $\lambda(p, Q)$ as in (80), the derivative $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ satisfies:*

$$\dot{\rho}(\dot{h}_0)(p) = 2\lambda(p, P_0)' \nabla F \left(\int m(x, \theta^*(p)) dP_0(x) \right) \int m(x, \theta^*(p)) \dot{h}_0(x) p_0(x) d\mu(x),$$

where $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ is Borel measurable and satisfies $\theta^*(p) \in \Xi(p, P_0)$ (as in (73)) for all $p \in \mathbb{S}^{d_\theta}$.

Proof: The existence of a Borel measurable $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ satisfying $\theta^*(p) \in \Xi(p, P_0)$ for all $p \in \mathbb{S}^{d_\theta}$ follows from Lemma B.11. Moreover, notice that indeed $\dot{\rho}(\dot{h}_0) \in \mathcal{C}(\mathbb{S}^{d_\theta})$ for all $\dot{h}_0 \in \dot{\mathbf{P}}$ as implied by Lemmas B.12 and B.13. We next establish that $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ is a continuous linear operator and then verify it is indeed the derivative of $\rho : \mathbf{P} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$. Linearity is immediate, while continuity follows by noting that by the Cauchy-Schwarz inequality:

$$\begin{aligned} & \sup_{\|\dot{h}_0\|_{L_\mu^2} = 1} \|\dot{\rho}(\dot{h}_0)\|_\infty \\ & \leq \sup_{\|\dot{h}_0\|_{L_\mu^2} = 1} \sup_{p \in \mathbb{S}^{d_\theta}} \{2\|\lambda(p, P_0)\| \times \sup_{v \in V} \|\nabla F(v)\|_F \times \sqrt{d_m} \sup_{(x, \theta) \in \mathcal{X} \times \Theta} \|m(x, \theta)\| \times \|\dot{h}_0\|_{L_\mu^2} \times \|p_0\|_{L_\mu^2}\} < \infty, \end{aligned} \quad (106)$$

where the final inequality is implied by Lemma B.9, Assumptions 3.2(i), 3.3(ii)-(iii) and $\|p_0\|_{L_\mu^2} = 1$.

In order to show $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ is the derivative of $\rho : \mathbf{P} \rightarrow \mathbf{B}$ at p_0 we need to establish that:

$$\lim_{\eta_0 \rightarrow 0} \int_{\mathbb{S}^{d_\theta}} \left\{ \frac{\nu(p, \Theta_0(P_{\eta_0})) - \nu(p, \Theta_0(P_0))}{\eta_0} - \dot{\rho}(\dot{h}_0)(p) \right\} dB(p) = 0 \quad (107)$$

for all curves $\eta \mapsto h_\eta$ in \mathbf{P} with $h_0 = p_0$ and all finite Borel measures B . However, by the mean value theorem:

$$\begin{aligned} \lim_{\eta_0 \rightarrow 0} \int_{\mathbb{S}^{d_\theta}} \frac{\nu(p, \Theta_0(P_{\eta_0})) - \nu(p, \Theta_0(P_0))}{\eta_0} dB(p) &= \lim_{\eta_0 \rightarrow 0} \int_{\mathbb{S}^{d_\theta}} \frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta)) \Big|_{\eta = \bar{\eta}(p, \eta_0)} dB(p) \\ &= \int_{\mathbb{S}^{d_\theta}} \frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta)) \Big|_{\eta=0} dB(p) = \int_{\mathbb{S}^{d_\theta}} \dot{\rho}(\dot{h}_0)(p) dB(p), \end{aligned} \quad (108)$$

where the first equality holds at each p for some $\bar{\eta}(p, \eta_0)$ a convex combination of η_0 and 0. The second equality in turn follows by Lemma B.13 justifying the use of the dominated convergence theorem, while the final equality follows by Lemma B.12 and the definition of $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$. Therefore, from (108), (107) is established. ■

Proof of Theorem 3.2: We employ the framework in Chapter 5.2 in Bickel et al. (1993) following their notation (except our \mathbf{P} corresponds to their \mathbf{S}). Let $\mathbf{B} \equiv \mathcal{C}(\mathbb{S}^{d_\theta})$ and \mathbf{B}^* denote the set of finite Borel measures on \mathbb{S}^{d_θ} , which by Corollary 14.15 in Aliprantis and Border (2006) is the dual space of \mathbf{B} . Let \mathbf{P} be as in Theorems B.1 and B.2, $p_0 \equiv \sqrt{dP_0/d\mu}$ and $\rho : \mathbf{P} \rightarrow \mathbf{B}$ be pointwise given by $\rho(h_\eta) \equiv \nu(\cdot, \Theta_0(P_\eta))$ where $h_\eta = \sqrt{dP_\eta/d\mu}$. By Theorem B.2, $\rho : \mathbf{P} \rightarrow \mathbf{B}$ has pathwise weak derivative $\dot{\rho}$ at p_0 . For $p \mapsto \theta^*(p)$ as in Lemma B.11 and any $B \in \mathbf{B}^*$ then let:

$$\dot{\rho}^T(B)(x) \equiv \int_{\mathbb{S}^{d_\theta}} 2\lambda(p, P_0)' H(\theta^*(p)) \{m(x, \theta^*(p)) - E[m(X, \theta^*(p))]\} p_0(x) dB(p). \quad (109)$$

We first show that $\dot{\rho}^T : \mathbf{B}^* \rightarrow \dot{\mathbf{P}}$ is the adjoint of $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathbf{B}$. Towards this end we establish that (i) $\dot{\rho}^T(B)$ is well defined for any $B \in \mathbf{B}^*$, (ii) $\dot{\rho}^T(B) \in \dot{\mathbf{P}}$ and finally (iii) $\dot{\rho}^T$ is the adjoint of ρ .

By Assumption 3.2(ii), Lemma B.11 and Lemmas 4.51 and 4.52 in Aliprantis and Border (2006) the function

$(x, p) \mapsto m(x, \theta^*(p))$ is jointly measurable and hence so is $p \mapsto E[m(X, \theta^*(p))]$. Similarly, $p \mapsto H(\theta^*(p))$ is measurable by continuity of $\theta \mapsto H(\theta)$ (see (79)) and Lemma B.11, while $p \mapsto \lambda(p, P_0)$ and $x \mapsto p_0(x)$ are trivially measurable by Lemma B.10 and $p_0 \in L^2_\mu$. The joint measurability of $(p, x) \mapsto (\lambda(p, P_0), H(\theta^*(p)), m(x, \theta^*(p)), E[m(X, \theta^*(p))], p_0(x))$ in $\mathbf{R}^{d_F} \times \mathbf{R}^{d_F \times d_m} \times \mathbf{R}^{d_m} \times \mathbf{R}^{d_m} \times \mathbf{R}$ then follows from Lemma 4.49 in Aliprantis and Border (2006) and hence:

$$(p, x) \mapsto 2\lambda(p, P_0)' H(\theta^*(p)) \{m(x, \theta^*(p)) - E[m(X, \theta^*(p))]\} p_0(x) \quad (110)$$

is jointly measurable by continuity of the composition. We conclude $\dot{\rho}^T(B)$ is a well defined measurable function for all $B \in \mathbf{B}^*$. Moreover, for $|B|$ the total variation of the measure B , Lemma B.9 and $\int p_0^2(x) d\mu(x) = 1$ imply:

$$\int_{\mathcal{X}} (\dot{\rho}^T(B)(x))^2 d\mu(x) \leq \sup_{p \in \mathbb{S}^{d_\theta}} 16 \|\lambda(p, P_0)\|^2 \times \sup_{v \in V} \|\nabla F(v)\|_F^2 \times \sup_{(x, \theta) \in \mathcal{X} \times \Theta} \|m(x, \theta)\|^2 \times |B|(\mathbb{S}^{d_\theta}) < \infty, \quad (111)$$

which verifies $\dot{\rho}^T(B) \in L^2_\mu$ for all $B \in \mathbf{B}^*$. Similarly, since $p_0^2 = dP_0/d\mu$, exchanging the order of integration yields:

$$\int_{\mathcal{X}} \dot{\rho}^T(B)(x) p_0(x) d\mu(x) = 2 \int_{\mathcal{X}} \int_{\mathbb{S}^{d_\theta}} \lambda(p, P_0)' H(\theta^*(p)) \{m(x, \theta^*(p)) - E[m(X, \theta^*(p))]\} dB(p) dP_0(x) = 0. \quad (112)$$

Therefore, by Theorem B.1 and (112) we conclude $\dot{\rho}^T(B) \in \dot{\mathbf{P}}$ for all $B \in \mathbf{B}^*$. In addition, we note that since

$$\int_{\mathbb{S}^{d_\theta}} \dot{\rho}(h)(p) dB(p) = \int_{\mathcal{X}} h(x) \dot{\rho}^T(B)(x) d\mu(x) \quad (113)$$

by Theorem B.1 implying $\int h(x) p(x) d\mu(x) = 0$ for any $h \in \dot{\mathbf{P}}$, we conclude $\dot{\rho}^T : \mathbf{B}^* \rightarrow \dot{\mathbf{P}}$ is the adjoint of $\dot{\rho} : \dot{\mathbf{P}} \rightarrow \mathbf{B}^*$.

Since $\dot{\mathbf{P}}_0$ is linear by Theorem B.1, Theorem B.2 and Theorem 5.2.1 in Bickel et al. (1993) establishes that:

$$\begin{aligned} \text{Cov} \left(\int_{\mathbb{S}^{d_\theta}} \mathbb{G}(p) dB_1(p), \int_{\mathbb{S}^{d_\theta}} \mathbb{G}(p) dB_2(p) \right) &= \frac{1}{4} \int_{\mathcal{X}} \dot{\rho}^T(B_1)(x) \dot{\rho}^T(B_2)(x) d\mu(x) \\ &= \int_{\mathbb{S}^{d_\theta}} \int_{\mathbb{S}^{d_\theta}} \lambda(p, P_0)' H(\theta^*(p)) \Omega(\theta^*(p), \theta^*(q)) H(\theta^*(q))' \lambda(q, P_0) dB_1(p) dB_2(q) \end{aligned} \quad (114)$$

for any $B_1, B_2 \in \mathbf{B}^*$, with the second equality following from $p_0^2 = dP_0/d\mu$ and reversing the order of integration. Letting B_1 and B_2 equal the degenerate probability measures at p_1 and p_2 in (114) then concludes the proof. ■

APPENDIX C - Proof of Theorems 4.1, 4.2, 4.3 and 4.4

Lemma C.1. *Let $\{W_i, X_i\}_{i=1}^n$ be an i.i.d. sample with $W_i \in \mathbf{R}$ independent of X_i and $E[W_i^2] < \infty$ and define $\mathcal{F} \equiv \{f : \mathcal{X} \times \mathbf{R} \rightarrow \mathbf{R} : f(x, w) = wm(x, \theta), \theta \in \Theta\}$. If Assumptions 3.1(i) and 3.2(ii) hold, then \mathcal{F} is Donsker.*

Proof: Let $m^{(i)}(x, \theta)$ denote the i^{th} component of the vector $m(x, \theta) \in \mathbf{R}^{d_m}$. For any $\theta_1, \theta_2 \in \Theta$ we then obtain:

$$\sup_{x \in \mathcal{X}} |w(m^{(i)}(x, \theta_1) - m^{(i)}(x, \theta_2))| \leq \sup_{(x, \theta) \in \mathcal{X} \times \Theta} \|\nabla_\theta m(x, \theta)\|_F \times \|\theta_1 - \theta_2\| \times |w| = G(w) \|\theta_1 - \theta_2\|, \quad (115)$$

where the inequality is implied by the Cauchy-Schwarz inequality and the mean value theorem, while the equality holds for $G(w) \equiv M|w|$ for some constant M due to Assumption 3.2(ii). It follows that the class \mathcal{F} is Lipschitz in $\theta \in \Theta$ and therefore by Theorem 2.7.11 in van der Vaart and Wellner (1996) we conclude that:

$$N_{[\cdot]}(2\epsilon \|G\|_{L^2}, \mathcal{F}, \|\cdot\|_{L^2}) \leq N(\epsilon, \Theta, \|\cdot\|). \quad (116)$$

Letting $D = \text{diam}(\Theta)$ and $u = \epsilon/2 \|G\|_{L^2}$, a change of variables and result (116) then allow us to conclude that:

$$\begin{aligned} \int_0^\infty \sqrt{\log N_{[\cdot]}(\epsilon, \mathcal{F}, \|\cdot\|_{L^2})} d\epsilon &= 2 \|G\|_{L^2} \int_0^\infty \sqrt{\log N_{[\cdot]}(2u \|G\|_{L^2}, \mathcal{F}, \|\cdot\|_{L^2})} du \\ &\leq 2 \|G\|_{L^2} \int_0^\infty \sqrt{N(u, \Theta, \|\cdot\|)} du = 2 \|G\|_{L^2} \int_0^D \sqrt{d_\theta \log(D/u)} du < \infty, \end{aligned} \quad (117)$$

where the final equality holds due to $N(u, \Theta, \|\cdot\|) \leq (\text{diam}(\Theta)/u)^{d_\theta}$. Since $\|G\|_{L^2}^2 = ME[W_i^2] < \infty$, the claim of the Lemma then follows from result (117) and Theorem 2.5.6 in van der Vaart and Wellner (1996). ■

Lemma C.2. For any neighborhood $N(P_0) \subseteq \mathbf{M}$ there is a convex neighborhood $N'(P_0) \subseteq \mathbf{M}$ with $N'(P_0) \subseteq N(P_0)$.

Proof: Let \mathbf{M}_s denote the set of signed, finite, countably additive Borel measures on \mathcal{X} endowed with the τ -topology. Note that $\mathbf{M} \subset \mathbf{M}_s$ and that \mathbf{M}_s is a topological vector space. For \mathcal{F} the set of bounded scalar valued measurable functions on \mathcal{X} and every $(f, \nu) \in \mathcal{F} \times \mathbf{M}_s$ define $p_f : \mathbf{M}_s \rightarrow \mathbf{R}$ by $p_f(\nu) = |\int f d\nu|$. The set of functionals $\{p_f\}_{f \in \mathcal{F}}$ is then a family of seminorms on \mathbf{M}_s that, by Lemma 5.76(2) in Aliprantis and Border (2006), generates the τ -topology. Therefore, Theorem 5.73 in Aliprantis and Border (2006) establishes that (\mathbf{M}_s, τ) is a locally convex topological vector space. Moreover, by Lemma 2.53 in Aliprantis and Border (2006), the τ -topology in \mathbf{M} is the relative topology on \mathbf{M} induced by (\mathbf{M}_s, τ) . Hence, letting $N^o(P_0)$ denote the interior of $N(P_0)$ (relative to \mathbf{M}), we obtain that $N^o(P_0) = N_s(P_0) \cap \mathbf{M}$ for some open set $N_s(P_0) \subseteq \mathbf{M}_s$. However, since (\mathbf{M}_s, τ) is locally convex, there exists an open (in \mathbf{M}_s) convex neighborhood of P_0 with $N'_s(P_0) \subseteq N_s(P_0)$. Defining $N'(P_0) = N'_s(P_0) \cap \mathbf{M}$ we obtained the desired result by convexity of \mathbf{M} . ■

Lemma C.3. Let Assumptions 3.1, 3.2, 3.3 and 3.4 hold, for any $Q \in \mathbf{M}$ define $Q_\tau \equiv \tau Q + (1 - \tau)P_0$ and let $\Xi(p, Q)$ be as in (73). Then, there is a $N(P_0) \subseteq \mathbf{M}$ such that for all $Q \in N(P_0)$, $p \in \mathbb{S}^{d_\theta}$ and $\tau_0 \in (0, 1)$:

$$\left. \frac{\partial}{\partial \tau} \nu(p, \Theta_0(Q_\tau)) \right|_{\tau=\tau_0} = \lambda(p, Q_{\tau_0})' \nabla F \left(\int m(x, \theta^*) dQ_{\tau_0}(x) \int m(x, \theta^*) (dQ(x) - dP_0(x)) \right) \quad \text{for any } \theta^* \in \Xi(p, Q_{\tau_0}).$$

Proof: First observe that by Lemma C.2 we may without loss of generality assume neighborhoods are convex. Hence, if $Q \in N(P_0)$, then $Q_\tau \in N(P_0)$ for all $\tau \in [0, 1]$. Since $\tau \mapsto F(\int m(x, \theta) dQ_\tau(x))$ is continuously differentiable in τ in a neighborhood of P_0 by Assumption 3.3(ii)-(iii), Lemma B.7 and Corollary 5 in Milgrom and Segal (2002) imply that on a neighborhood of P_0 the function $\tau \mapsto \nu(p, \Theta_0(Q_\tau))$ is directionally differentiable at any $\tau_0 \in (0, 1)$ with:

$$\left. \frac{\partial}{\partial \tau_+} \nu(p, \Theta_0(Q_\tau)) \right|_{\tau=\tau_0} = \max_{\theta^* \in \Xi(p, Q_{\tau_0})} \lambda(p, Q_{\tau_0})' \nabla F \left(\int m(x, \theta^*) dQ_{\tau_0}(x) \int m(x, \theta^*) (dQ(x) - dP_0(x)) \right) \quad (118)$$

$$\left. \frac{\partial}{\partial \tau_-} \nu(p, \Theta_0(Q_\tau)) \right|_{\tau=\tau_0} = \min_{\theta^* \in \Xi(p, Q_{\tau_0})} \lambda(p, Q_{\tau_0})' \nabla F \left(\int m(x, \theta^*) dQ_{\tau_0}(x) \int m(x, \theta^*) (dQ(x) - dP_0(x)) \right) \quad (119)$$

where $\frac{\partial}{\partial \tau_+}$ and $\frac{\partial}{\partial \tau_-}$ denote the right and left derivatives. Hence, if $p \in \mathbb{S}^{d_\theta}$ is such that $\Xi(p, Q_{\tau_0})$ is a singleton, then the Lemma follows. On the other hand, if $p \in \mathbb{S}^{d_\theta}$ is such that $\Xi(p, Q_{\tau_0})$ is not a singleton, then by Lemma B.8 there is $\theta_L \in \Xi(p, Q_{\tau_0})$ such that $F^{(i)}(\int m_S(x, \theta_L) dQ_{\tau_0}(x)) < 0$ for $1 \leq i \leq d_{F_S}$. For $\lambda_L(p, Q_{\tau_0})$ the Lagrange multiplier of the linear constraints, the complementary slackness conditions imply that for any $\theta^* \in \Xi(p, Q_{\tau_0})$:

$$\begin{aligned} \lambda(p, Q_{\tau_0})' \nabla F \left(\int m(x, \theta^*) dQ_{\tau_0}(x) \int m(x, \theta^*) (dQ(x) - dP_0(x)) \right) \\ = -\lambda_L(p, Q_{\tau_0})' \nabla F_L \left(\int m_L(x) dQ_{\tau_0}(x) \int m_L(x) (dQ(x) - dP_0(x)) \right). \end{aligned} \quad (120)$$

It follows that (118) and (119) also agree when $\Xi(p, Q_{\tau_0})$ is not a singleton and the Lemma is established. ■

Lemma C.4. Let $N(P_0) \subseteq \mathbf{M}$ be a neighborhood of P_0 and $\Gamma : \mathbb{S}^{d_\theta} \times N(P_0) \rightarrow \mathbf{R}^k$ be upper hemicontinuous. If $W \subseteq \mathbb{S}^{d_\theta}$ is compact and $\Gamma(p, P_0)$ is a singleton $\forall p \in W$, then for every $\epsilon > 0$ there is a $N'(P_0) \subseteq N(P_0)$ such that:

$$\sup_{p \in W} \sup_{Q \in N'(P_0)} \sup_{\gamma \in \Gamma(p, Q)} \|\gamma - \Gamma(p, P_0)\| < \epsilon.$$

Proof: Since $\Gamma(p, P_0)$ is singleton valued for all $p \in W$ and $\Gamma : \mathbb{S}^{d_\theta} \times N(P_0) \rightarrow \mathbf{R}^k$ is upper hemicontinuous, compactness of W implies $p \mapsto \Gamma(p, P_0)$ is a uniformly continuous function on W . Hence, there is a $\delta_1 > 0$ with:

$$\sup_{p, \tilde{p} \in W : \|p - \tilde{p}\| < \delta_1} \|\Gamma(p, P_0) - \Gamma(\tilde{p}, P_0)\| < \frac{\epsilon}{2}. \quad (121)$$

For any $\delta > 0$ and $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P_0)$ let $\Gamma^\delta(p, Q) \equiv \{\gamma \in \mathbf{R}^k : \inf_{\tilde{\gamma} \in \Gamma(p, Q)} \|\gamma - \tilde{\gamma}\| < \delta\}$. Since the correspondence $\Gamma : \mathbb{S}^{d_\theta} \times N(P_0) \rightarrow \mathbf{R}^k$ is upper hemicontinuous, it follows that for each $p \in \mathbb{S}^{d_\theta}$ and any $\delta_2 < \epsilon/2$ there is a neighborhood $S_{\delta_2}(p, P_0) \times M_{\delta_2}(p, P_0)$ of (p, P_0) in $\mathbb{S}^{d_\theta} \times \mathbf{M}$ such that $\Gamma(\tilde{p}, Q) \subseteq \Gamma^{\delta_2}(p, P_0)$ for all $(\tilde{p}, Q) \in S_{\delta_2}(p, P_0) \times M_{\delta_2}(p, P_0)$.

Letting $N_{\delta_1}(p)$ denote an open ball of radius δ_1 around p , notice that $\{N_{\delta_1}(p) \cap S_{\delta_2}(p, P_0)\}_{p \in \mathbb{S}^{d_\theta}}$ covers W . Therefore, by compactness, there exists a finite set $\{p_i\}_{i=1}^K$ such that $\{N_{\delta_1}(p_i) \cap S_{\delta_2}(p_i, P_0)\}_{i=1}^K$ is a subcover for W . Now let $N'(P_0) \equiv \bigcap_{i=1}^K M_{\delta_2}(p_i, P_0)$, which is a neighborhood of P_0 due to $K < \infty$. Then, every $(p, Q) \in W \times N'(P_0)$ must satisfy $(p, Q) \in N_{\delta_1}(p_i) \cap S_{\delta_2}(p_i, P_0) \times M_{\delta_2}(p_i, P_0)$ for some $1 \leq i \leq K$ and hence we obtain:

$$\begin{aligned} & \sup_{(p, Q) \in W \times N'(P_0)} \sup_{\gamma \in \Gamma(p, Q)} \|\gamma - \Gamma(p, P_0)\| \\ & \leq \max_{1 \leq i \leq K} \left\{ \sup_{p \in W \cap N_{\delta_1}(p_i)} \|\Gamma(p_i, P_0) - \Gamma(p, P_0)\| + \sup_{\gamma \in \Gamma^{\delta_2}(p_i, P_0)} \|\gamma - \Gamma(p_i, P_0)\| \right\} < \epsilon \end{aligned} \quad (122)$$

where the final inequality follows from (121) and $\delta_2 < \epsilon/2$. Hence, the claim of the Lemma follows from (122). ■

Lemma C.5. *Let Assumption 4.1 hold and P_* denote inner probability. Then for every neighborhood $N(P_0) \subseteq \mathbf{M}$:*

$$\liminf_{n \rightarrow \infty} P_*(\hat{P}_n \in N(P_0)) = 1 .$$

Proof: The empirical measure \hat{P}_n is not measurable in \mathbf{M} with respect to the Borel σ -field generated by the τ -topology, which is why we employ inner probabilities; see Chapter 6.2 in Dembo and Zeitouni (1998). Let \mathcal{F} denote the set of scalar bounded measurable functions on \mathcal{X} and for every $(f, \nu) \in \mathcal{F} \times \mathbf{M}$ define $p_f : \mathbf{M} \rightarrow \mathbf{R}$ by $p_f(\nu) \equiv \int f(x) d\nu(x)$. Since the τ -topology is the coarsest topology making $\nu \mapsto p_f(\nu)$ continuous for all $f \in \mathcal{F}$, it follows that for arbitrary but finite K , $\{W_i\}_{i=1}^K$ open sets in \mathbf{R} , and $\{f_i\}_{i=1}^K \in \mathcal{F}$, the sets of the form:

$$\bigcap_{i=1}^K \{Q \in \mathbf{M} : p_{f_i}(Q) \in W_i\} \quad (123)$$

constitute a base for the τ -topology. Thus, since P_0 is in the interior of $N(P_0)$, there exists an integer K_0 , a finite collection $\{f_i\}_{i=1}^{K_0}$ and an $\epsilon > 0$ such that $\bigcap_{i=1}^{K_0} \{Q \in \mathbf{M} : |\int f_i(x)(dP_0(x) - dQ(x))| \leq \epsilon\} \subseteq N(P_0)$. Hence,

$$\liminf_{n \rightarrow \infty} P_*(\hat{P}_n \in N(P_0)) \geq \liminf_{n \rightarrow \infty} P(\max_{1 \leq i \leq k_0} |\int f_i(x)(d\hat{P}_n(x) - dP_0(x))| \leq \epsilon) = 1 , \quad (124)$$

where the final equality follows from the law of large numbers since each f_i is bounded. ■

Lemma C.6. *Let $\{W_i, X_i\}_{i=1}^n$ be an i.i.d. sample with $W_i \in \mathbf{R}$ independent of X_i and $E[W_i^2] < \infty$. Further define $\hat{P}_{n, \tau} \equiv \tau \hat{P}_n + (1 - \tau)P_0$ for any $\tau \in [0, 1]$ and let $\Xi(p, Q)$ be as in (73). If Assumptions 3.1, 3.2, 3.3, 3.4 and 4.1 hold and P_0^W and \hat{P}_n^W are the joint population and empirical measures of (X_i, W_i) respectively, then:*

$$\sup_{p \in \mathbb{S}^{d_\theta}} \sup_{\tau \in [0, 1]} \sup_{\theta \in \Xi(p, \hat{P}_{n, \tau})} |\sqrt{n} \lambda(p, P_0)' H(\theta^*(p)) \int w(m(x, \theta) - m(x, \theta^*(p)))(d\hat{P}_n^W(x, w) - dP_0^W(x, w))| = o_p(1) ,$$

where $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ is a Borel measurable mapping that satisfies $\theta^*(p) \in \Xi(p, P_0)$ for all $p \in \mathbb{S}^{d_\theta}$.

Proof: If $N(P_0) \subseteq \mathbf{M}$ is convex and $\hat{P}_n \in N(P_0)$, then $\hat{P}_{n, \tau} \in N(P_0)$ for all $\tau \in [0, 1]$. Therefore, by Lemmas B.4, C.2 and C.5 we obtain that with inner probability tending to one $\Xi(p, \hat{P}_{n, \tau})$ is well defined for all $(p, \tau) \in \mathbb{S}^{d_\theta} \times [0, 1]$. Also let $\lambda(p, P_0) = (\lambda_S(p, P_0)', \lambda_L(p, P_0)')'$ with $\lambda_S(p, P_0)$ and $\lambda_L(p, P_0)$ the Lagrange multipliers for the strictly convex and linear constraints respectively. Similarly partition the matrix $H(\theta) = (H_S(\theta)', H_L(\theta)')$ and define:

$$\gamma(p) \equiv H_S(\theta^*(p))' \lambda_S(p, P_0) \quad \Delta_n(p, \theta) \equiv \int w(m_S(x, \theta) - m_S(x, \theta^*(p)))(d\hat{P}_n^W(x, w) - dP_0^W(x, w)) . \quad (125)$$

Then notice that since $\int A\theta(d\hat{P}_n^W(x, w) - dP_0(x, w)) = 0$ for all $\theta \in \Theta$, we obtain from definition (125) that:

$$\sqrt{n} \lambda(p, P_0)' H(\theta^*(p)) \int w(m(x, \theta) - m(x, \theta^*(p)))(d\hat{P}_n^W(x, w) - dP_0^W(x, w)) = \sqrt{n} \gamma(p)' \Delta_n(p, \theta) . \quad (126)$$

We establish the Lemma by studying the process $\sqrt{n} \gamma(p)' \Delta_n(p, \theta)$ on the following three subsets of \mathbb{S}^{d_θ} :

$$\mathbb{S}_L^{d_\theta} \equiv \{p \in \mathbb{S}^{d_\theta} : \Xi(p, P_0) \text{ not a singleton}\} \quad \mathbb{S}_L^{d_\theta, \delta} \equiv \{p \in \mathbb{S}^{d_\theta} : \inf_{\tilde{p} \in \mathbb{S}_L^{d_\theta}} \|p - \tilde{p}\| < \delta\} \quad (\mathbb{S}_L^{d_\theta, \delta})^c \equiv \mathbb{S}^{d_\theta} \setminus \mathbb{S}_L^{d_\theta, \delta} . \quad (127)$$

By Lemma B.8 and the complementary slackness conditions $\lambda_S(p, P_0) = 0 \forall p \in \mathbb{S}_L^{d_\theta}$. Since $p \mapsto \lambda_S(p, P_0)$ is continuous by Lemmas B.7 and B.10, \mathbb{S}^{d_θ} compact implies $p \mapsto \lambda_S(p, P_0)$ is uniformly continuous. By Assumption 3.3(ii)-(iii), $\|H_S(\theta^*(p))\|_F \leq \sup_{v \in V} \|\nabla F(v)\|_F < \infty$. Therefore, for every $M > 0$ we may find a $\delta > 0$ such that:

$$\sup_{p \in \mathbb{S}_L^{d_\theta, \delta}} \|\gamma(p)\| \leq \sup_{v \in V} \|\nabla F(v)\|_F \times \sup_{p \in \mathbb{S}_L^{d_\theta, \delta}} \|\lambda_S(p, P_0)\| < \frac{1}{M}. \quad (128)$$

Hence, since $\sup_{\theta \in \Theta} \|\int w m_S(x, \theta)(d\hat{P}_n^W(x, w) - dP_0^W(x, w))\| = O_p(n^{-\frac{1}{2}})$ by Lemma C.1, it follows from the Cauchy-Schwarz inequality and result (128) that for any $\epsilon, \eta > 0$ there exists a $\delta_1 > 0$ such that:

$$\begin{aligned} \limsup_{n \rightarrow \infty} P\left(\sup_{p \in \mathbb{S}_L^{d_\theta, \delta_1}} \sup_{\tau \in [0, 1]} \sup_{\theta \in \Xi(p, \hat{P}_{n, \tau})} |\sqrt{n}\gamma(p)' \Delta_n(p, \theta)| > \epsilon\right) \\ \leq \limsup_{n \rightarrow \infty} P\left(\sup_{p \in \mathbb{S}_L^{d_\theta, \delta_1}} \|\gamma(p)\| \times \sup_{\theta \in \Theta} 2\|\sqrt{n} \int w m_S(x, \theta)(d\hat{P}_n^W(x, w) - dP_0^W(x, w))\| > \epsilon\right) < \frac{\eta}{2}. \end{aligned} \quad (129)$$

Since $p \mapsto \|\gamma(p)\|$ is bounded by Lemma B.9 and Assumptions 3.3(ii)-(iii), Lemma C.1 implies for any $\epsilon, \eta > 0$:

$$\limsup_{n \rightarrow \infty} P\left(\sup_{p \in \mathbb{S}^{d_\theta}} \|\gamma(p)\| \times \sup_{\|\theta_1 - \theta_2\| < \delta_2} \|\sqrt{n} \int w(m_S(x, \theta_1) - m_S(x, \theta_2))(d\hat{P}_n^W(x, w) - dP_0^W(x, w))\| > \epsilon\right) < \frac{\eta}{2} \quad (130)$$

for some $\delta_2 > 0$. The correspondence $p \mapsto \Xi(p, P_0)$ is upper hemicontinuous by Lemma B.4 and singleton valued on $(\mathbb{S}_L^{d_\theta, \delta_1})^c$. Hence, applying Lemma C.5 and C.4 with $W = (\mathbb{S}_L^{d_\theta, \delta_1})^c$ and $\Gamma(p, Q) = \Xi(p, Q)$ we obtain:

$$\sup_{p \in (\mathbb{S}_L^{d_\theta, \delta_1})^c} \sup_{\tau \in [0, 1]} \sup_{\theta \in \Xi(p, \hat{P}_{n, \tau})} \|\theta - \theta^*(p)\| = o_p(1). \quad (131)$$

Therefore, using the Cauchy-Schwarz inequality and combining the results in (130) and (131) we finally obtain:

$$\begin{aligned} \limsup_{n \rightarrow \infty} P\left(\sup_{p \in (\mathbb{S}_L^{d_\theta, \delta_1})^c} \sup_{\tau \in [0, 1]} \sup_{\theta \in \Xi(p, \hat{P}_{n, \tau})} |\sqrt{n}\gamma(p) \Delta_n(p, \theta)| > \epsilon\right) \\ < \frac{\eta}{2} + \limsup_{n \rightarrow \infty} P\left(\sup_{p \in (\mathbb{S}_L^{d_\theta, \delta_1})^c} \sup_{\tau \in [0, 1]} \sup_{\theta \in \Xi(p, \hat{P}_{n, \tau})} \|\theta - \theta^*(p)\| > \delta_2\right) = \frac{\eta}{2}. \end{aligned} \quad (132)$$

Since $\epsilon, \eta > 0$ were arbitrary, the claim of the Lemma then follows from (126), (129) and (132). ■

Lemma C.7. Define $\hat{P}_{n, \tau} \equiv \tau \hat{P}_n + (1 - \tau)P_0$ and let Assumptions 3.1, 3.2, 3.3, 3.4 and 4.1 hold. Then:

$$\sup_{p \in \mathbb{S}^{d_\theta}} \sup_{\tau \in [0, 1]} \sup_{\theta \in \Xi(p, \hat{P}_{n, \tau})} \|\lambda(p, \hat{P}_{n, \tau})' \nabla F\left(\int m(x, \theta) d\hat{P}_{n, \tau}(x)\right) - \lambda(p, P_0)' \nabla F\left(\int m(x, \theta^*(p)) dP_0(x)\right)\| = o_p(1),$$

where $\Xi(p, Q)$ is as in (73) and $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ is a Borel measurable function with $\theta^*(p) \in \Xi(p, P_0)$ for all $p \in \mathbb{S}^{d_\theta}$.

Proof: Fix $\epsilon > 0$ and note that by Lemmas B.7 and B.10 there exists a neighborhood $N(P_0)$ in \mathbf{M} such that the correspondence $(p, Q) \mapsto \lambda(p, Q)$ is upper hemicontinuous and singleton valued for all $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P_0)$. Applying Lemmas C.2 and C.4 with $W = \mathbb{S}^{d_\theta}$ then implies that there exists a convex neighborhood $N'(P_0) \subseteq \mathbf{M}$ such that:

$$\sup_{p \in \mathbb{S}^{d_\theta}} \sup_{Q \in N'(P_0)} \|\lambda(p, Q) - \lambda(p, P_0)\| < \epsilon. \quad (133)$$

Since $N'(P_0)$ is convex, $\hat{P}_n \in N'(P_0)$ implies $\hat{P}_{n, \tau} \in N'(P_0)$ for all $\tau \in [0, 1]$. Therefore, it follows that:

$$\liminf_{n \rightarrow \infty} P\left(\sup_{p \in \mathbb{S}^{d_\theta}} \sup_{\tau \in [0, 1]} \|\lambda(p, \hat{P}_{n, \tau}) - \lambda(p, P_0)\| < \epsilon\right) \geq \liminf_{n \rightarrow \infty} P(\hat{P}_n \in N(P_0)) = 1, \quad (134)$$

where the final equality follows from Lemma C.5 and probabilities should be interpreted as inner probabilities.

Next, note that by Assumptions 3.3(iii) and Lemmas C.5 and C.2 it follows that for V as in Assumption 3.3(i):

$$\liminf_{n \rightarrow \infty} P\left(\int m(x, \theta) d\hat{P}_{n, \tau}(x) \in V \text{ for all } (\theta, \tau) \in \Theta \times [0, 1]\right) = 1. \quad (135)$$

Hence, since $\sup_{v \in V} \|\nabla F(v)\|_F < \infty$ by Assumption 3.3(ii), we obtain from results (133), (134) and (135) that:

$$\sup_{p \in \mathbb{S}^{d_\theta}} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \|(\lambda(p, \hat{P}_{n,\tau}) - \lambda(p, P_0))' \nabla F(\int m(x, \theta) d\hat{P}_{n,\tau}(x))\| = o_p(1). \quad (136)$$

Let $\lambda(p, P_0) = (\lambda_S(p, P_0)', \lambda_L(p, P_0)')$ with $\lambda_S(p, P_0)$ and $\lambda_L(p, P_0)$ the Lagrange multipliers corresponding the strictly convex and linear constraints respectively. By Assumption 3.2(i) and the law of large numbers we obtain:

$$\sup_{\tau \in [0,1]} \left\| \int m_S(x) (d\hat{P}_{n,\tau}(x) - dP_0(x)) \right\| \leq \left\| \int m_S(x) (d\hat{P}_n(x) - dP_0(x)) \right\| = o_p(1). \quad (137)$$

Hence, since $v \mapsto \nabla F(v)$ is continuous on V by Assumption 3.3(ii), we conclude from Lemma B.9 and result (135):

$$\begin{aligned} & \sup_{p \in \mathbb{S}^{d_\theta}} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \left\| \lambda_L(p, P_0)' (\nabla F_L(\int m_L(x) d\hat{P}_{n,\tau}(x)) - \nabla F_L(\int m_L(x) dP_0(x))) \right\| \\ & \leq \sup_{p \in \mathbb{S}^{d_\theta}} \|\lambda_L(p, P_0)\| \times \sup_{\tau \in [0,1]} \left\| \nabla F_L(\int m_L(x) d\hat{P}_{n,\tau}(x)) - \nabla F_L(\int m_L(x) dP_0(x)) \right\| = o_p(1). \end{aligned} \quad (138)$$

For the strictly convex constraints, let $\mathbb{S}_L^{d_\theta}$, $\mathbb{S}_L^{d_\theta, \delta}$ and $(\mathbb{S}_L^{d_\theta, \delta})^c$ be as in (127). For any $\epsilon > 0$, there is a $\delta_1 > 0$ with:

$$\begin{aligned} & \limsup_{n \rightarrow \infty} P\left(\sup_{p \in \mathbb{S}_L^{d_\theta, \delta_1}} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \left| \lambda_S(p, P_0)' (\nabla F_S(\int m_S(x, \theta) d\hat{P}_{n,\tau}(x)) - \nabla F_S(\int m_S(x, \theta^*(p)) dP_0(x))) \right| > \epsilon \right) \\ & \leq \limsup_{n \rightarrow \infty} P\left(\sup_{p \in \mathbb{S}_L^{d_\theta, \delta_1}} \sup_{(\tau, \theta) \in [0,1] \times \Theta} 2|\lambda_S(p, P_0)' \nabla F_S(\int m_S(x, \theta) d\hat{P}_{n,\tau}(x))| > \epsilon \right) = 0, \end{aligned} \quad (139)$$

where the final equality follows from (128) and (135). Next, note $\theta \mapsto E[m(X, \theta)]$ is uniformly continuous due to Assumptions 3.2(i)-(ii), the dominated convergence theorem and compactness of Θ . Hence, by Lemma C.1 and (131):

$$\begin{aligned} & \sup_{p \in (\mathbb{S}_L^{d_\theta, \delta_1})^c} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \left\| \int m(x, \theta) d\hat{P}_{n,\tau}(x) - \int m(x, \theta^*(p)) dP_0(x) \right\| \\ & \leq \sup_{p \in (\mathbb{S}_L^{d_\theta, \delta_1})^c} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \left\| \int (m(x, \theta) - m(x, \theta^*(p))) dP_0(x) \right\| + o_p(1) = o_p(1). \end{aligned} \quad (140)$$

Therefore, since $v \mapsto \nabla F(v)$ is uniformly continuous on V by Assumption 3.3(ii), (135), (140) and Lemma B.9 imply:

$$\sup_{p \in (\mathbb{S}_L^{d_\theta, \delta_1})^c} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \left| \lambda_S(p, P_0)' (\nabla F_S(\int m_S(x, \theta) d\hat{P}_{n,\tau}(x)) - \nabla F_S(\int m_S(x, \theta^*(p)) dP_0(x))) \right| = o_p(1). \quad (141)$$

The claim of the Lemma then follows by direct calculation and results (136), (138), (139) and (141). ■

Lemma C.8. *Let Assumptions 3.1, 3.2, 3.3, 3.4 and 4.1 hold and $\Xi(p, P_0)$ be as in (73). It then follows that:*

$$\sup_{p \in \mathbb{S}^{d_\theta}} |\sqrt{n} \{(\nu(p, \Theta_0(\hat{P}_n)) - \nu(p, \Theta_0(P_0))) - \lambda(p, P_0)' H(\theta^*(p)) \int m(x, \theta^*(p)) (d\hat{P}_n(x) - dP_0(x))\}| = o_p(1),$$

where $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ is a Borel measurable mapping satisfying $\theta^*(p) \in \Xi(p, P_0)$ for all $p \in \mathbb{S}^{d_\theta}$.

Proof: For every $\tau \in [0, 1]$ define $\hat{P}_{n,\tau} \equiv \tau \hat{P}_n + (1 - \tau)P_0$ and notice that $\hat{P}_{n,0} = P_0$ and $\hat{P}_{n,1} = \hat{P}_n$. Employing the mean value theorem, which is valid by Lemmas C.2, C.3 and C.5, we can then conclude that uniformly in $p \in \mathbb{S}^{d_\theta}$:

$$\begin{aligned} & \nu(p, \Theta_0(\hat{P}_n)) - \nu(p, \Theta_0(P_0)) \\ & = \lambda(p, \hat{P}_{n,\tau_0(p)})' \nabla F(\int m(x, \tilde{\theta}(p)) d\hat{P}_{n,\tau_0(p)}(x)) \int m(x, \tilde{\theta}(p)) (d\hat{P}_n(x) - dP_0(x)) + o_p(n^{-\frac{1}{2}}) \end{aligned} \quad (142)$$

for some $\tau_0 : \mathbb{S}^{d_\theta} \rightarrow (0, 1)$ and $\tilde{\theta} : \mathbb{S}^{d_\theta} \rightarrow \Theta$ is such that $\tilde{\theta}(p) \in \Xi(p, \hat{P}_{n,\tau_0(p)})$ for all $p \in \mathbb{S}^{d_\theta}$. In turn, by Lemma C.1:

$$\sup_{p \in \mathbb{S}^{d_\theta}} \left\| \sqrt{n} \int m(x, \tilde{\theta}(p)) (d\hat{P}_n(x) - dP_0(x)) \right\| \leq \sup_{\theta \in \Theta} \left\| \sqrt{n} \int m(x, \theta) (d\hat{P}_n(x) - dP_0(x)) \right\| = O_p(1). \quad (143)$$

Therefore, exploiting Lemma C.7 together with results (142) and (143) we obtain that uniformly in $p \in \mathbb{S}^{d_\theta}$:

$$\begin{aligned} \sqrt{n}(\nu(p, \Theta_0(\hat{P}_n)) - \nu(p, \Theta_0(P_0))) &= \sqrt{n}\lambda(p, P_0)'H(\theta^*(p)) \int m(x, \tilde{\theta}(p))(d\hat{P}_n(x) - dP_0(x)) + o_p(1) \\ &= \sqrt{n}\lambda(p, P_0)'H(\theta^*(p)) \int m(x, \theta^*(p))(d\hat{P}_n(x) - dP_0(x)) + o_p(1), \end{aligned} \quad (144)$$

where the second equality follows from Lemma C.6 applied with W_i equal to one with probability one. ■

Lemma C.9. *Let Assumptions 3.1-3.4 and 4.1 hold, $\Xi(p, P_0)$ be as in (73) and $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ with $\theta^*(p) \in \Xi(p, P_0) \forall p \in \mathbb{S}^{d_\theta}$. Then $\mathcal{F} \equiv \{f : \mathcal{X} \rightarrow \mathbf{R} : f(x) = \lambda(p, P_0)'H(\theta^*(p))m(x, \theta^*(p)) \text{ for some } p \in \mathbb{S}^{d_\theta}\}$ is Donsker (in $\mathcal{C}(\mathbb{S}^{d_\theta})$).*

Proof: For notational simplicity, let us define a stochastic process G_n defined on \mathbb{S}^{d_θ} to be pointwise given by:

$$G_n(p) \equiv \sqrt{n}\lambda(p, P_0)'H(\theta^*(p)) \int m(x, \theta^*(p))(d\hat{P}_n(x) - dP_0(x)). \quad (145)$$

We first show convergence of the marginals. If $p \in \mathbb{S}^{d_\theta}$ is such that $\Xi(p, P_0)$ is a singleton, then $m : \mathcal{X} \times \Theta \rightarrow \mathbf{R}^{d_m}$, $v \mapsto \nabla F(v)$ and $p \mapsto \lambda(p, P_0)$ being bounded by Assumptions 3.2(i), 3.3(ii)-(iii) and Lemma B.9 then imply:

$$\sqrt{n}\lambda(p, P_0)'H(\theta^*(p)) \int m(x, \theta^*(p))(d\hat{P}_n(x) - dP_0(x)) \xrightarrow{L} N(0, \sigma^2(p)), \quad (146)$$

where $\sigma^2(p) \equiv \text{Var}(\lambda(p, P_0)'H(\theta^*(p))m(x, \theta^*(p)))$. If $\Xi(p, P_0)$ is not a singleton, then by Lemma B.8 there is a $\theta_L \in \Xi(p, P_0)$ such that $F^{(i)}(\int m_S(x, \theta_L)dP_0(x)) < 0$ for $1 \leq i \leq d_{F_S}$. For $\lambda(p, P_0) = (\lambda_S(p, P_0)', \lambda_L(p, P_0)')$ with $\lambda_S(p, P_0)$, $\lambda_L(p, P_0)$ corresponding to the strictly convex and linear constraints respectively, the complementary slackness conditions imply $\lambda_S(p, P_0) = 0$ and $\lambda_L(p, P_0)'(A\theta - F_L(E[m_L(X)])) = 0$ for any $\theta \in \Xi(p, P_0)$. Hence,

$$\begin{aligned} \sqrt{n}\lambda(p, P_0)'H(\theta^*(p)) \int m(x, \theta)(d\hat{P}_n(x) - dP_0(x)) \\ = -\sqrt{n}\lambda_L(p, P_0)'\nabla F_L(E[m_L(X)]) \int m_L(x)(d\hat{P}_n(x) - dP_0(x)) \xrightarrow{L} N(0, \sigma^2(p)). \end{aligned} \quad (147)$$

Hence, for all $p \in \mathbb{S}^{d_\theta}$ the limiting law of $G_n(p)$ is normal and independent of how $\theta^*(p) \in \Xi(p, P_0)$ is selected.

Next, note that in (146)-(147) it was argued $p \mapsto G_n(p)$ is bounded, while identical arguments to those in (100)-(102) show $p \mapsto G_n(p)$ is continuous with probability one. Hence, $G_n \in \mathcal{C}(\mathbb{S}^{d_\theta})$ almost surely and to establish the Lemma we only need to show the asymptotic uniform equicontinuity of G_n . Towards this end, we consider the linear and strictly convex moment inequalities separately. For the linear inequalities we obtain by Assumption 3.3(iii):

$$\begin{aligned} \sup_{\|p_1 - p_2\| < \delta} \sqrt{n}|(\lambda_L(p_1, P_0) - \lambda_L(p_2, P_0))'\nabla F_L(E[m_L(X)]) \int m_L(x)(d\hat{P}_n(x) - dP_0(x))| \\ \leq \sup_{\|p_1 - p_2\| < \delta} \|\lambda_L(p_1, P_0) - \lambda_L(p_2, P_0)\| \times \sup_{v \in V} \|\nabla F(v)\|_F \times \|\sqrt{n} \int m_L(x)(d\hat{P}_n(x) - dP_0(x))\|. \end{aligned} \quad (148)$$

However, $p \mapsto \lambda_L(p, P_0)$ is uniformly continuous by Lemmas B.7, B.10 and \mathbb{S}^{d_θ} compact. Hence, Assumption 3.3(ii) and $\sqrt{n} \int m_L(x)(d\hat{P}_n(x) - dP_0(x)) = O_p(1)$ imply that for any $\eta, \epsilon > 0$ there is a δ_L such that:

$$\limsup_{n \rightarrow \infty} P\left(\sup_{\|p_1 - p_2\| < \delta_L} \|\lambda_L(p_1, P_0) - \lambda_L(p_2, P_0)\| \times \sup_{v \in V} \|\nabla F(v)\|_F \times \|\sqrt{n} \int m_L(x)(d\hat{P}_n(x) - dP_0(x))\| > \epsilon\right) < \eta. \quad (149)$$

The asymptotic uniform equicontinuity of the term corresponding to the linear constraints follows from (148)-(149).

For examining the strictly convex inequalities partition $H(\theta) = (H_S(\theta)', H_L(\theta)')$ and define:

$$C_n(p) \equiv \gamma(p)'\Delta_n(\theta^*(p)) \quad \gamma(p) \equiv H_S(\theta^*(p))'\lambda_S(p, P_0) \quad \Delta_n(\theta) \equiv \sqrt{n} \int m_S(x, \theta)(d\hat{P}_n(x) - dP_0(x)). \quad (150)$$

Let $\mathbb{S}_L^{d_\theta}$, $\mathbb{S}_L^{d_\theta, \delta}$ and $(\mathbb{S}_L^{d_\theta, \delta})^c$ be as in (127) and recall that as shown in (129), we can obtain that for any $\eta, \epsilon > 0$:

$$\limsup_{n \rightarrow \infty} P\left(\sup_{p_1, p_2 \in \mathbb{S}_L^{d_\theta, \delta_1}, \|p_1 - p_2\| < \delta} |C_n(p_1) - C_n(p_2)| > 2\epsilon\right) \leq \limsup_{n \rightarrow \infty} P\left(\sup_{p \in \mathbb{S}_L^{d_\theta, \delta_1}} \|\gamma(p)\| \times \sup_{\theta \in \Theta} \|\Delta_n(\theta)\| > \epsilon\right) < 2\eta \quad (151)$$

for some $\delta_1 > 0$. On the other hand, by the Cauchy-Schwarz inequality it also follows that for any $p_1, p_2 \in \mathbb{S}^{d_\theta}$:

$$|C_n(p_1) - C_n(p_2)| \leq \|\gamma(p_1) - \gamma(p_2)\| \times \sup_{\theta \in \Theta} \|\Delta_n(\theta)\| + \sup_{p \in \mathbb{S}^{d_\theta}} \|\gamma(p)\| \times \|\Delta_n(\theta^*(p_1)) - \Delta_n(\theta^*(p_2))\|. \quad (152)$$

By construction, $p \mapsto \Xi(p, P_0)$ is singleton-valued for all $p \in (S_L^{d_\theta, \frac{\delta_1}{2}})^c$. Hence, since $p \mapsto \Xi(p, P_0)$ is upper hemicontinuous on \mathbb{S}^{d_θ} by Lemma B.4 it follows that $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ is continuous on $(S_L^{d_\theta, \frac{\delta_1}{2}})^c$ and thus so is $p \mapsto H_S(\theta^*(p))$ by Assumptions 3.3(ii)-(iii), 3.2(i)-(ii) and the dominated convergence theorem. Since $p \mapsto \lambda_S(p, P_0)$ is continuous by Lemma B.10, by compactness $p \mapsto \gamma(p)$ is uniformly continuous on $(S_L^{d_\theta, \frac{\delta_1}{2}})^c$. Therefore, since by Lemma C.1 we have $\sup_{\theta \in \Theta} \|\Delta_n(\theta)\| = O_p(1)$, it follows that for any $\epsilon, \eta > 0$ there exists a $\delta_\gamma > 0$ such that:

$$\limsup_{n \rightarrow \infty} P\left(\sup_{p_1, p_2 \in (S_L^{d_\theta, \frac{\delta_1}{2}})^c, \|p_1 - p_2\| < \delta_\gamma} \|\gamma(p_1) - \gamma(p_2)\| \times \sup_{\theta \in \Theta} \|\Delta_n(\theta)\| > \epsilon\right) < \eta. \quad (153)$$

In addition, notice that as argued in (130) for ever $\epsilon, \eta > 0$ there also exists a $\delta_\theta > 0$ such that:

$$\limsup_{n \rightarrow \infty} P\left(\sup_{p \in \mathbb{S}^{d_\theta}} \|\lambda_S(p, P_0)\| \times \sup_{\|\theta_1 - \theta_2\| < \delta_\theta} \|\Delta_n(\theta_1) - \Delta_n(\theta_2)\| > \epsilon\right) < \eta. \quad (154)$$

However, as argued $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ is continuous on $(S_L^{d_\theta, \frac{\delta_1}{2}})^c$ and hence uniformly continuous by compactness. Therefore, for δ_θ as in (154), there exists a $\delta_p > 0$ so that $\|\theta^*(p_1) - \theta^*(p_2)\| < \delta_\theta$ for all $p_1, p_2 \in (S_L^{d_\theta, \frac{\delta_1}{2}})^c$ with $\|p_1 - p_2\| < \delta_p$. In lieu of (154), this in turn allows us to conclude that:

$$\limsup_{n \rightarrow \infty} P\left(\sup_{p \in \mathbb{S}^{d_\theta}} \|\gamma(p)\| \times \sup_{\|p_1 - p_2\| < \delta_p, p_1, p_2 \in (S_L^{d_\theta, \frac{\delta_1}{2}})^c} \|\Delta_n(\theta^*(p_1)) - \Delta_n(\theta^*(p_2))\| > \epsilon\right) < \eta. \quad (155)$$

Therefore, since $a + b > 2\epsilon$ implies $a > \epsilon$ or $b > \epsilon$, we obtain from results (152), (153) and (155) that we must have:

$$\limsup_{n \rightarrow \infty} P\left(\sup_{\|p_1 - p_2\| < \delta_2, p_1, p_2 \in (S_L^{d_\theta, \frac{\delta_1}{2}})^c} |C_n(p_1) - C_n(p_2)| > 2\epsilon\right) < 2\eta \quad (156)$$

if $\delta_2 < \min\{\delta_\gamma, \delta_p\}$. Moreover, notice that for any $\delta_S < \min\{\frac{\delta_1}{2}, \delta_2\}$ and $p_1, p_2 \in \mathbb{S}^{d_\theta}$ with $\|p_1 - p_2\| < \delta_S$ it follows that if $p_1 \in (S_L^{d_\theta, \delta_1})^c$, then $p_1, p_2 \in (S_L^{d_\theta, \frac{\delta_1}{2}})^c$. We therefore obtain:

$$\sup_{\|p_1 - p_2\| < \delta_S} |C_n(p_1) - C_n(p_2)| \leq \sup_{\substack{p_1, p_2 \in \mathbb{S}^{d_\theta, \delta_1} \\ \|p_1 - p_2\| < \delta_S}} |C_n(p_1) - C_n(p_2)| + \sup_{\substack{p_1, p_2 \in (S_L^{d_\theta, \frac{\delta_1}{2}})^c \\ \|p_1 - p_2\| < \delta_S}} |C_n(p_1) - C_n(p_2)|. \quad (157)$$

Hence, using inequality (157), results (151) and (156) together with arguments as in (156) we finally obtain:

$$\limsup_{n \rightarrow \infty} P\left(\sup_{\|p_1 - p_2\| < \delta_S} |C_n(p_1) - C_n(p_2)| > 4\epsilon\right) < 4\eta \quad (158)$$

which establishes the asymptotic uniform equicontinuity of C_n . Since the asymptotic uniform equicontinuity of the term corresponding to linear constraints was established in (149) we conclude G_n is asymptotically uniform equicontinuous as well. In turn, because \mathbb{S}^{d_θ} is totally bounded under $\|\cdot\|$, the process G_n is asymptotically tight in $\mathcal{C}(\mathbb{S}^{d_\theta})$ by Theorem 1.5.7 in van der Vaart and Wellner (1996). The Lemma then follows from the convergence of the marginals and Theorem 1.5.4, Addendum 1.5.8 and Theorem 1.3.10 in van der Vaart and Wellner (1996). ■

Proof of Theorem 4.1: By Lemma C.8, $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ is asymptotically linear with influence function $\psi : \mathcal{X} \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$:

$$\psi(x) \equiv \lambda(\cdot, P_0)' H(\theta^*(\cdot)) \{m(x, \theta^*(\cdot)) - E[m(X, \theta^*(\cdot))]\} \quad (159)$$

where $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ with $\theta^*(p) \in \Xi(p, P_0)$; which establishes (ii). By Theorem 3.2, $x \mapsto \psi(x)$ is the efficient influence function, and hence regularity of $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ follows from Lemma C.9 and Theorem 18.1 in Kosorok (2008); which establishes (i). The stated convergence in distribution is then immediate from Lemmas C.8 and C.9, while the limiting process having the efficient covariance kernel is a direct result of the characterization of $I^{-1}(p_1, p_2)$ obtained

in Theorem 3.2; which establishes (iii). ■

Proof of Theorem 4.2: First notice that since $\{K_n\}$ is convex and compact valued, Corollary 1.10 in Li et al. (2002) implies that $d_H(K_n, \Theta_0(P_0)) = \|\nu(\cdot, K_n) - \nu(\cdot, \Theta_0(P_0))\|_\infty$. Therefore, since by hypothesis $p \mapsto \nu(p, K_n)$ is a regular estimator for $p \mapsto \nu(p, \Theta_0(P_0))$, we obtain from Theorems B.1, B.2 and Proposition 5.2.1 in Bickel et al. (1993) that:

$$\liminf_{n \rightarrow \infty} E[L(\sqrt{n}d_H(K_n, \Theta_0(P_0)))] = \liminf_{n \rightarrow \infty} E[L(\sqrt{n}\|\nu(\cdot, K_n) - \nu(\cdot, \Theta_0(P_0))\|_\infty)] \geq E[L(\|\mathbb{G}_0\|_\infty)] . \quad (160)$$

Next, we aim to show that $E[L(\sqrt{n}d_H(\hat{\Theta}_n, \Theta_0))]$ attains the lower bound. Towards this end, define:

$$G_n(p) \equiv \sqrt{n}(\nu(p, \Theta_0(\hat{P}_n)) - \nu(p, \Theta_0(P_0))) , \quad (161)$$

and note $G_n \in \mathcal{C}(\mathbb{S}^{d_\theta})$ almost surely. Define the functional $f \mapsto L(\|f\|_\infty) \wedge C$, which is continuous and bounded on $\mathcal{C}(\mathbb{S}^{d_\theta})$ for any constant $C > 0$. Therefore, by Theorem 4.1 and the Portmanteau Theorem we can conclude that:

$$\limsup_{C \uparrow \infty} \limsup_{n \rightarrow \infty} |E[L(\|G_n\|_\infty) \wedge C] - E[L(\|\mathbb{G}_0\|_\infty) \wedge C]| = 0 . \quad (162)$$

By Theorem 4.1, Addendum 1.5.8 in van der Vaart and Wellner (1996) and compactness of \mathbb{S}^{d_θ} , the sample paths of \mathbb{G}_0 are bounded almost surely. Since $L(a) \leq Ma^\kappa$ by hypothesis, Proposition A.2.3 in van der Vaart and Wellner (1996) yields $E[L(\|\mathbb{G}_0\|_\infty)] \leq ME[\|\mathbb{G}_0\|_\infty^\kappa] < \infty$ and hence by the monotone convergence theorem we conclude:

$$\limsup_{C \uparrow \infty} |E[L(\|\mathbb{G}_0\|_\infty)] - E[L(\|\mathbb{G}_0\|_\infty) \wedge C]| = 0 . \quad (163)$$

By Assumption 3.3(ii)-(iii) and Lemmas B.9 and C.2 there exists a convex neighborhood $N(P_0) \subseteq \mathbf{M}$ such that (i) $\nabla F(\int m(x, \theta)dQ(x))$ is uniformly bounded in $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P_0)$; (ii) $p \mapsto \lambda(p, Q)$ is uniformly bounded on $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P_0)$ and (iii) the conditions of Lemma C.3 are satisfied for all $Q \in N(P_0)$. For every $\tau \in [0, 1]$ define $\hat{P}_{n,\tau} \equiv \tau \hat{P}_{n,\tau} + (1 - \tau)P_0$ and note that if $\hat{P}_n \in N(P_0)$ then (142) holds so that uniformly in $p \in \mathbb{S}^{d_\theta}$:

$$G_n = \Delta_n \quad \Delta_n \equiv p \mapsto \lambda(p, \hat{P}_{n,\tau_0(p)})' \nabla F \left(\int m(x, \tilde{\theta}(p)) d\hat{P}_{n,\tau_0(p)}(x) \right) \int \sqrt{nm}(x, \tilde{\theta}(p))(d\hat{P}_n(x) - dP_0(x)) , \quad (164)$$

for some $\tau_0 : \mathbb{S}^{d_\theta} \rightarrow (0, 1)$ and $\tilde{\theta} : \mathbb{S}^{d_\theta} \rightarrow \Theta$ with $\tilde{\theta}(p) \in \Xi(p, \hat{P}_{n,\tau_0(p)})$ for $\Xi(p, Q)$ as in (73) (and set $\Delta_n = 0$ if $\hat{P}_n \notin N(P_0)$). By compactness of Θ , definition of $N(P_0)$ and $m(x, \theta)$ bounded by Assumption 3.2(i) we must have

$$\max\{\|G_n\|_\infty, \|\Delta_n\|_\infty\} \leq \sqrt{n}C_0 , \quad (165)$$

for some $C_0 > 0$. Therefore, since $L(a) \leq Ma^\kappa$ for all $a \in \mathbf{R}_+$, (164) holding if $\hat{P}_n \in N(P_0)$ and (165) yield:

$$\limsup_{n \rightarrow \infty} |E[L(\|G_n\|_\infty)] - E[L(\|\Delta_n\|_\infty)]| \leq \limsup_{n \rightarrow \infty} 2MC_0^\kappa n^{\frac{\kappa}{2}} P(\hat{P}_n \notin N(P_0)) . \quad (166)$$

However, as shown in (124), there exists a finite collection $\{f_j\}_{j=1}^{k_0}$ of bounded functions and an $\epsilon > 0$ such that $\{Q \in \mathbf{M} : \max_{1 \leq j \leq k_0} |\int f_j(x)(dQ(x) - dP_0(x))| \leq \epsilon\} \subseteq N(P_0)$. Therefore, (166) and Bernstein's inequality imply:

$$\limsup_{n \rightarrow \infty} |E[L(\|G_n\|_\infty)] - E[L(\|\Delta_n\|_\infty)]| \leq 2MC_0^\kappa \limsup_{n \rightarrow \infty} \sum_{j=1}^{k_0} n^{\frac{\kappa}{2}} P(|\int f_j(x)(d\hat{P}_n(x) - dP_0(x))| > \epsilon) = 0 . \quad (167)$$

From result (167) and applying Cauchy-Schwarz and Markov's inequalities we can then conclude that:

$$\begin{aligned} \limsup_{n \rightarrow \infty} |E[L(\|G_n\|_\infty)] - E[L(\|G_n\|_\infty) \wedge C]| &= \limsup_{n \rightarrow \infty} |E[L(\|\Delta_n\|_\infty)] - E[L(\|\Delta_n\|_\infty) \wedge C]| \\ &\leq \limsup_{n \rightarrow \infty} E[L(\|\Delta_n\|_\infty) 1\{L(\|\Delta_n\|_\infty) > C\}] \leq \limsup_{n \rightarrow \infty} \frac{1}{C} E[L^2(\|\Delta_n\|_\infty)] . \end{aligned} \quad (168)$$

By construction of $N(P_0)$, there exists a compact set $\mathbf{C} \subset \mathbf{R}^{d_m}$ such that $\lambda(p, Q)' \nabla F(\int m(x, \theta)dQ(x)) \in \mathbf{C}$ for all $(p, \theta, Q) \in \mathbb{S}^{d_\theta} \times \Theta \times N(P_0)$. Let $\mathcal{G} \equiv \{g : \mathcal{X} \rightarrow \mathbf{R} : g(x) = c' m(x, \theta) \text{ for some } (c, \theta) \in \mathbf{C} \times \Theta\}$ and note that by Assumption 3.2(i)-(ii) and compactness of \mathbf{C} , there exists a $C_1 > 0$ such that $\sup_{x \in \mathcal{X}} |g(x)| \leq C_1$ for all $g \in \mathcal{G}$.

Moreover, for any $(c_1, \theta_1) \in \mathbf{C} \times \Theta$ and $(c_2, \theta_2) \in \mathbf{C} \times \Theta$ we also obtain by Assumption 3.2(i)-(ii) that:

$$\begin{aligned} & \sup_{x \in \mathcal{X}} |c'_1 m(x, \theta_1) - c'_2 m(x, \theta_2)| \\ & \leq \left\{ \sup_{(x, \theta) \in \mathcal{X} \times \Theta} \|m(x, \theta)\| + \sup_{(x, \theta) \in \mathcal{X} \times \Theta} \|\nabla_{\theta} m(x, \theta)\|_F \times \sup_{c \in \mathbf{C}} \|c\| \right\} \times \{\|c_1 - c_2\| + \|\theta_1 - \theta_2\|\}, \end{aligned} \quad (169)$$

and hence the class \mathcal{G} is Lipschitz in $(\theta, c) \in \Theta \times \mathbf{C}$. Letting $\|\cdot\| + \|\cdot\|$ denote the sum of the Euclidean norms on $\mathbf{R}^{d_{\theta}}$ and \mathbf{R}^{d_m} , we then obtain by Theorem 2.7.11 in van der Vaart and Wellner (1996), that:

$$N_{[]} (2\epsilon C_1, \mathcal{G}, \|\cdot\|_{\infty}) \leq N(\epsilon, \Theta \times \mathbf{C}, \|\cdot\| + \|\cdot\|) \lesssim \epsilon^{-(d_m + d_{\theta})}. \quad (170)$$

Consequently, since $\Delta_n = 0$ whenever $\hat{P}_n \notin N(P_0)$, result (168) and $L(a) \leq Ma^{\kappa}$ for all $a \in \mathbf{R}_+$ then imply that:

$$\begin{aligned} \limsup_{n \rightarrow \infty} E[L^2(\|\Delta_n\|_{\infty})] & \leq \limsup_{n \rightarrow \infty} M^2 E[\|\Delta_n\|_{\infty}^{2\kappa}] \\ & \leq \limsup_{n \rightarrow \infty} M^2 E[\sup_{g \in \mathcal{G}} |\frac{1}{\sqrt{n}} \sum_{i=1}^n \{g(X_i) - E[g(X_i)]\}|^{2\kappa}] \lesssim \left(\int_0^1 \sqrt{1 + \log N_{[]}(\epsilon C_1, \mathcal{G}, \|\cdot\|_{\infty})} d\epsilon \right)^{2\kappa} < \infty \end{aligned} \quad (171)$$

where the third inequality follows from Theorem 2.14.1 in van der Vaart and Wellner (1996) and the last result is immediate from (170). Combining results (168) and (171), we can finally obtain:

$$\limsup_{C \uparrow \infty} \limsup_{n \rightarrow \infty} |E[L(\|G_n\|_{\infty})] - E[L(\|G_n\|_{\infty}) \wedge C]| \leq \limsup_{C \uparrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{C} E[L^2(\|\Delta_n\|_{\infty})] = 0. \quad (172)$$

Since $\sqrt{nd}_H(\hat{\Theta}_n, \Theta_0(P_0)) = \|G_n\|_{\infty}$ by Corollary 1.10 in Li et al. (2002), (162), (163) and (172) imply:

$$\limsup_{n \rightarrow \infty} E[L(\sqrt{nd}_H(\hat{\Theta}_n, \Theta_0(P_0)))] - E[L(\|G_0\|_{\infty})] = \limsup_{n \rightarrow \infty} |E[L(\|G_n\|_{\infty})] - E[L(\|G_0\|_{\infty})]| = 0, \quad (173)$$

which together with result (160) establishes the claim of the Theorem. ■

Proof of Theorem 4.3: For any metric space $(\mathbb{D}, \|\cdot\|_{\mathbb{D}})$ let $BL_M(\mathbb{D})$ denote the set of Lipschitz real functions on \mathbb{D} whose absolute value and Lipschitz constant are bounded by M . To establish the Theorem, it then suffices to show:

$$\sup_{f \in BL_1(\mathcal{C}(\mathbb{S}^{d_{\theta}}))} |E[f(G_n^*)|\{X_i\}_{i=1}^n] - E[f(G_0)]| = o_p(1), \quad (174)$$

due to Theorem 1.12.4 in van der Vaart and Wellner (1996). Towards this end, note that Lemma C.1 implies that:

$$\begin{aligned} & \sup_{p \in \mathbb{S}^{d_{\theta}}} \|\sqrt{n} \int w \{m(x, \hat{\theta}(p)) - \int m(x, \hat{\theta}(p)) d\hat{P}_n(x)\} d\hat{P}_n^W(x, w)\| \\ & \leq \sup_{\theta \in \Theta} \|\sqrt{n} \int w m(x, \theta) d\hat{P}_n^W(x, w)\| + \sup_{(x, \theta) \in (\mathcal{X} \times \Theta)} \|m(x, \theta)\| \times |\sqrt{n} \int w d\hat{P}_n^W(x, w)| = O_p(1) \end{aligned} \quad (175)$$

due to $W_i \perp X_i$, $E[W_i] = 0$ by Assumption 4.2(ii) and $(x, \theta) \mapsto m(x, \theta)$ being uniformly bounded by Assumption 3.2(i). Next, let $\lambda_S(p, P_0)$ and $\nabla F_S(v)$ be the elements of $\lambda(p, P_0)$ and $\nabla F(v)$ corresponding to the strictly convex constraints. By result (140) together with (125), (128) and $(x, \theta) \mapsto m(x, \theta)$ bounded we then obtain:

$$\sup_{p \in \mathbb{S}^{d_{\theta}}} |\lambda_S(p, P_0)' \nabla F_S \left(\int m(x, \theta^*(p)) dP_0(x) \right) \left(\int m(x, \hat{\theta}(p)) d\hat{P}_n(x) - \int m(x, \theta^*(p)) dP_0(x) \right)| = o_p(1). \quad (176)$$

The analogous result for the linear constraints is immediate by the law of large numbers. Therefore, by results (175) and (176), $E[W_i] = 0$ by Assumption 4.2(ii) and Lemmas C.6 and C.7 we obtain:

$$\begin{aligned} & \sqrt{n} \lambda(p, \hat{P}_n)' \nabla F \left(\int m(x, \hat{\theta}(p)) d\hat{P}_n(x) \right) \int w \{m(x, \hat{\theta}(p)) - \int m(x, \hat{\theta}(p)) d\hat{P}_n(x)\} d\hat{P}_n^W(x, w) \\ & = \sqrt{n} \lambda(p, P_0)' \nabla F \left(\int m(x, \theta^*(p)) dP_0(x) \right) \int w \{m(x, \hat{\theta}(p)) - \int m(x, \hat{\theta}(p)) d\hat{P}_n(x)\} d\hat{P}_n^W(x, w) + o_p(1) \\ & = \sqrt{n} \lambda(p, P_0)' \nabla F \left(\int m(x, \theta^*(p)) dP_0(x) \right) \int w \{m(x, \theta^*(p)) - \int m(x, \theta^*(p)) dP_0(x)\} d\hat{P}_n^W(x, w) + o_p(1), \end{aligned} \quad (177)$$

where $\theta^* : \mathbb{S}^{d_\theta} \rightarrow \Theta$ satisfies $\theta^*(p) \in \Xi(p, P_0)$ (as in (73)) and (177) holds uniformly in $p \in \mathbb{S}^{d_\theta}$. Next, define:

$$\bar{G}_n^*(p) \equiv \sqrt{n} \lambda(p, P_0)' H(\theta^*(p)) \int w \{m(x, \theta^*(p)) - \int m(x, \theta^*(p)) dP_0(x)\} d\hat{P}_n^W(x, w), \quad (178)$$

and note that by Assumptions 3.2(i), 3.3(ii)-(iii) and Lemma B.9, $\bar{G}_n^* \in \mathcal{C}(\mathbb{S}^{d_\theta})$ almost surely. Since all $f \in BL_1(\mathcal{C}(\mathbb{S}^{d_\theta}))$ are bounded and have Lipschitz constant less than or equal to one, for any $\eta > 0$ we must have:

$$\sup_{f \in BL_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} |E[f(\bar{G}_n^*) - f(G_n^*) | \{X_i\}_{i=1}^n]| \leq \eta P(\|\bar{G}_n^* - G_n^*\|_\infty \leq \eta | \{X_i\}_{i=1}^n) + 2P(\|\bar{G}_n^* - G_n^*\|_\infty > \eta | \{X_i\}_{i=1}^n). \quad (179)$$

However, from (177), it follows that $P(\|\bar{G}_n^* - G_n^*\|_\infty > \eta | \{X_i\}_{i=1}^n) = o_p(1)$, and hence since η in (177) is arbitrary:

$$\sup_{f \in BL_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} |E[f(G_n^*) | \{X_i\}_{i=1}^n] - E[f(\bar{G}_n^*) | \{X_i\}_{i=1}^n]| = o_p(1). \quad (180)$$

To conclude, we note that by Lemma C.9 and Theorem 2.9.6 in van der Vaart and Wellner (1996), we have:

$$\sup_{f \in BL_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} |E[f(\bar{G}_n^*) | \{X_i\}_{i=1}^n] - E[f(G_0)]| = o_p(1), \quad (181)$$

and therefore the (180) and (181) verify (174) which establishes the claim of the Theorem. ■

Proof of Theorem 4.4: Let \bar{G}_n^* be defined as in (178) and note that by (177) $\|\bar{G}_n^* - G_n^*\|_\infty = o_p(1)$ unconditionally. Define a mapping $\Gamma : \mathcal{C}(\mathbb{S}^{d_\theta}) \rightarrow \mathcal{C}(\mathbb{S}^{d_\theta})$ pointwise by $\Gamma(f) = \Upsilon \circ f$. The continuous mapping theorem then yields:

$$|\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) - \sup_{p \in \hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p))| \leq \sup_{p \in \mathbb{S}^{d_\theta}} |\Upsilon(G_n^*(p)) - \Upsilon(\bar{G}_n^*(p))| = \|\Gamma(G_n^*) - \Gamma(\bar{G}_n^*)\|_\infty = o_p(1). \quad (182)$$

Next, let $\hat{p}^* \in \arg \max_{p \in \hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p))$ which is well defined by Assumption 4.3(ii) and continuity of $p \mapsto \bar{G}_n^*(p)$. Letting $\Pi_{\Psi_0} \hat{p}^*$ denote the projection of \hat{p}^* onto Ψ_0 and noting $\|\hat{p}^* - \Pi_{\Psi_0} \hat{p}^*\| \leq d_H(\hat{\Psi}_n, \Psi_0)$, we can then obtain:

$$\sup_{p \in \hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p)) - \sup_{p \in \Psi_0} \Upsilon(\bar{G}_n^*(p)) \leq \Upsilon(\bar{G}_n^*(\hat{p}^*)) - \Upsilon(\bar{G}_n^*(\Pi_{\Psi_0} \hat{p}^*)) \leq \sup_{\|p - \tilde{p}\| \leq d_H(\hat{\Psi}_n, \Psi_0)} |\Upsilon(\bar{G}_n^*(p)) - \Upsilon(\bar{G}_n^*(\tilde{p}))|. \quad (183)$$

Similarly, by analogous manipulations to the term $\sup_{p \in \Psi_0} \Upsilon(\bar{G}_n^*(p)) - \sup_{p \in \hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p))$, we can then conclude:

$$|\sup_{p \in \hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p)) - \sup_{p \in \Psi_0} \Upsilon(\bar{G}_n^*(p))| \leq \sup_{\|p - \tilde{p}\| \leq d_H(\hat{\Psi}_n, \Psi_0)} |\Upsilon(\bar{G}_n^*(p)) - \Upsilon(\bar{G}_n^*(\tilde{p}))|. \quad (184)$$

By Assumption 4.2(iii), Lemma C.9 and Theorem 2.9.2 in van der Vaart and Wellner (1996), $\bar{G}_n^* \xrightarrow{L} \bar{G}$ (unconditionally) for some tight Gaussian process \bar{G} in $\mathcal{C}(\mathbb{S}^{d_\theta})$. Therefore, it follows that $\sup_{p \in \mathbb{S}^{d_\theta}} |\bar{G}_n^*(p)|$ is asymptotically tight in \mathbf{R} . Fixing $\eta > 0$, $\epsilon > 0$ there then is a constant $K > 0$ such that:

$$\limsup_{n \rightarrow \infty} P(\sup_{p \in \mathbb{S}^{d_\theta}} |\bar{G}_n^*(p)| > K) < \eta. \quad (185)$$

By Assumption 4.3(i), $\Upsilon : \mathbf{R} \rightarrow \mathbf{R}$ is continuous and hence uniformly continuous on $[-K, K]$. Therefore, there is a $\delta_0 > 0$ such that $|\Upsilon(a_1) - \Upsilon(a_2)| < \epsilon$ whenever $|a_1 - a_2| < \delta_0$ with $a_1, a_2 \in [-K, K]$. Hence, we then obtain:

$$\begin{aligned} \limsup_{n \rightarrow \infty} P(\sup_{\|p - \tilde{p}\| \leq d_H(\hat{\Psi}_n, \Psi_0)} |\Upsilon(\bar{G}_n^*(p)) - \Upsilon(\bar{G}_n^*(\tilde{p}))| > \epsilon) \\ \leq \limsup_{n \rightarrow \infty} P(\sup_{\|p - \tilde{p}\| \leq d_H(\hat{\Psi}_n, \Psi_0)} |\bar{G}_n^*(p) - \bar{G}_n^*(\tilde{p})| > \delta_0) + \limsup_{n \rightarrow \infty} P(\sup_{p \in \mathbb{S}^{d_\theta}} |\bar{G}_n^*(p)| > K). \end{aligned} \quad (186)$$

Moreover, since the process $p \mapsto \bar{G}_n^*(p)$ is asymptotically tight in $\mathcal{C}(\mathbb{S}^{d_\theta})$ by Lemma 1.3.8 in van der Vaart and Wellner (1996), it then follows that there exists a $\gamma_0 > 0$ such that:

$$\begin{aligned} \limsup_{n \rightarrow \infty} P(\sup_{\|p - \tilde{p}\| \leq d_H(\hat{\Psi}_n, \Psi_0)} |\bar{G}_n^*(p) - \bar{G}_n^*(\tilde{p})| > \delta_0) \\ \leq \limsup_{n \rightarrow \infty} P(\sup_{\|p - \tilde{p}\| \leq \gamma_0} |\bar{G}_n^*(p) - \bar{G}_n^*(\tilde{p})| > \delta_0) + \limsup_{n \rightarrow \infty} P(d_H(\hat{\Psi}_n, \Psi_0) > \gamma_0) < \eta, \end{aligned} \quad (187)$$

due to $d_H(\hat{\Psi}_n, \Psi_0) = o_p(1)$ by hypothesis. Since ϵ, η where arbitrary, combining (182)-(186) we then obtain:

$$\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) = \sup_{p \in \Psi_0} \Upsilon(\bar{G}_n^*(p)) + o_p(1) . \quad (188)$$

Therefore, for $BL_1(\mathbf{R})$ as in (174), arguing as in (180) and using Theorem 4.3 and Theorem 10.8 in Kosorok (2008):

$$\begin{aligned} \sup_{f \in BL_1(\mathbf{R})} |E[f(\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) | \{X_i\}_{i=1}^n) - E[f(\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p)))]| \\ \leq \sup_{f \in BL_1(\mathbf{R})} |E[f(\sup_{p \in \Psi_0} \Upsilon(\bar{G}_n^*(p)) | \{X_i\}_{i=1}^n) - E[f(\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p)))]| + o_p(1) = o_p(1) . \end{aligned} \quad (189)$$

To conclude, observe that result (189) together with Lemma 10.11 in Kosorok (2008) imply that:

$$P(\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) \leq t | \{X_i\}_{i=1}^n) = P(\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p)) \leq t) + o_p(1) \quad (190)$$

for all $t \in \mathbf{R}$ that are continuity points of the cdf of $\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p))$. Moreover, since $c_{1-\alpha}$ is itself a continuity point, for any $\epsilon > 0$ there is an $\tilde{\epsilon} \leq \epsilon$ such that $c_{1-\alpha} \pm \tilde{\epsilon}$ are also continuity points and in addition:

$$P(\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p)) \leq c_{1-\alpha} - \tilde{\epsilon}) < 1 - \alpha < P(\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p)) \leq c_{1-\alpha} + \tilde{\epsilon}) , \quad (191)$$

due to the cdf of $\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p))$ being strictly increasing at $c_{1-\alpha}$. To conclude, define the event:

$$A_n \equiv \{P(\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) \leq c_{1-\alpha} - \tilde{\epsilon} | \{X_i\}_{i=1}^n) < 1 - \alpha < P(\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) \leq c_{1-\alpha} + \tilde{\epsilon} | \{X_i\}_{i=1}^n)\} \quad (192)$$

and observe that since $c_{1-\alpha} \pm \tilde{\epsilon}$ are continuity points of the cdf of $\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p))$, result (190) yields that:

$$\liminf_{n \rightarrow \infty} P(|\hat{c}_{1-\alpha} - c_{1-\alpha}| \leq \epsilon) \geq \liminf_{n \rightarrow \infty} P(A_n) = 1 , \quad (193)$$

which establishes the claim of the Theorem. ■

APPENDIX D - Proof of Theorems 5.1, 5.2, 5.3 and 5.4

Proof of Theorem 5.1: By Theorem 11.1 in Davydov et al. (1998), the cdf of $\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|_+$ is continuous and strictly increasing except possible at zero. However, since there is a $p_0 \in \mathbb{S}^{d_\theta}$ with $\text{Var}\{\mathbb{G}_0(p_0)\} > 0$, we have:

$$P(\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|_+ \leq 0) \leq P(\mathbb{G}_0(p_0) \leq 0) = 0.5 , \quad (194)$$

it follows that the median of $\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|_+$ is greater than or equal to zero. Therefore, $\alpha < 0.5$ guarantees that the cdf of $\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|_+$ is continuous and strictly increasing at $c_{1-\alpha}^{(1)}$. As a result, applying Theorem 4.4 with $\Upsilon(a) = |a|_+$ and $\Psi_0 = \hat{\Psi}_n = \mathbb{S}^{d_\theta}$ establishes that $\hat{c}_{1-\alpha}^{(1)} = c_{1-\alpha}^{(1)} + o_p(1)$.

Next observe that for two convex compact sets K_1, K_2 , we have $K_1 \subseteq K_2$ if and only if $\nu(p, K_1) \leq \nu(p, K_2)$ for all $p \in \mathbb{S}^{d_\theta}$. Hence, since $\nu(\cdot, K_1^\epsilon) = \nu(\cdot, K_1) + \epsilon$ for any convex compact K_1 and scalar $\epsilon > 0$, we obtain that:

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\Theta_0(P_0) \subseteq \hat{\Theta}_n^{\hat{c}_{1-\alpha}^{(1)}/\sqrt{n}}) &= \lim_{n \rightarrow \infty} P(\sup_{p \in \mathbb{S}^{d_\theta}} \sqrt{n}(\nu(p, \Theta_0(P_0)) - \nu(p, \Theta_0(\hat{P}_n))) \leq \hat{c}_{1-\alpha}^{(1)}) \\ &= \lim_{n \rightarrow \infty} P(\sup_{p \in \mathbb{S}^{d_\theta}} \sqrt{n}|\nu(p, \Theta_0(P_0)) - \nu(p, \Theta_0(\hat{P}_n))|_+ \leq \hat{c}_{1-\alpha}^{(1)}) = P(\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|_+ \leq c_{1-\alpha}) , \end{aligned} \quad (195)$$

where the second equality follows from $\hat{c}_{1-\alpha}^{(1)} \xrightarrow{P} c_{1-\alpha}^{(1)} > 0$ and the third follows from Slutsky, the continuous mapping theorem and $c_{1-\alpha}^{(1)}$ being a continuity point of the cdf of $\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|_+$. The claim of the Theorem is then established by (195), the definition of $c_{1-\alpha}^{(1)}$ and the cdf of $\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|_+$ being strictly increasing at $c_{1-\alpha}^{(1)}$. ■

Proof of Theorem 5.2: Since there exists a $p_0 \in \mathbb{S}^{d_\theta}$ such that $\text{Var}\{\mathbb{G}_0(p_0)\} > 0$, it follows by Theorem 11.1 in Davydov et al. (1998) that the cdf of $\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|$ is continuous and strictly increasing. Therefore, applying

Theorem 4.4 with $\Upsilon(a) = |a|$ and $\Psi_0 = \hat{\Psi}_n = \mathbb{S}^{d_\theta}$ implies $\hat{c}_{1-\alpha}^{(2)} = c_{1-\alpha}^{(2)} + o_p(1)$.

Next, observe that for any $\epsilon > 0$ and compact convex K such that $K^{-\epsilon} \neq \emptyset$ we have $\nu(\cdot, K^{-\epsilon}) = \nu(\cdot, K) - \epsilon$. Hence, since for convex compact sets K_1, K_2 , $K_1 \subseteq K_2$ if and only if $\nu(p, K_1) \leq \nu(p, K_2)$ for all $p \in \mathbb{S}^{d_\theta}$, we obtain:

$$\begin{aligned} K^{-\epsilon} \subseteq \Theta_0(P_0) \subseteq K^\epsilon &\iff \nu(p, K^{-\epsilon}) \leq \nu(p, \Theta_0(P_0)) \leq \nu(p, K^\epsilon) \quad \forall p \in \mathbb{S}^{d_\theta} \\ &\iff \sup_{p \in \mathbb{S}^{d_\theta}} |\nu(p, \Theta_0(P_0)) - \nu(p, K)| \leq \epsilon, \end{aligned} \quad (196)$$

where the second result holds provided $K^{-\epsilon} \neq \emptyset$. Moreover, since $\hat{c}_{1-\alpha}^{(2)} \xrightarrow{p} c_{1-\alpha}^{(2)} < \infty$ and $d_H(\hat{\Theta}_n, \Theta_0(P_0)) = o_p(1)$ by Theorem 4.2, Assumption 3.1(iii) implies that $\hat{\Theta}_n^{-\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}}$ is nonempty with probability tending to one. Hence, since $\hat{\Theta}_n$ is the convex hull of $\Theta_0(\hat{P}_n)$, it follows that $\nu(\cdot, \hat{\Theta}_n) = \nu(\cdot, \Theta_0(\hat{P}_n))$ and therefore (196) yields:

$$\lim_{n \rightarrow \infty} P(\hat{\Theta}_n^{-\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}} \subseteq \Theta_0(P_0) \subseteq \hat{\Theta}_n^{\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}}) = \lim_{n \rightarrow \infty} P(\sup_{p \in \mathbb{S}^{d_\theta}} \sqrt{n} |\nu(p, \Theta_0(\hat{P}_n)) - \nu(p, \Theta_0(P_0))| \leq \hat{c}_{1-\alpha}^{(2)}) = 1 - \alpha \quad (197)$$

where the final equality follows from $\hat{c}_{1-\alpha}^{(2)} \xrightarrow{p} c_{1-\alpha}^{(2)}$, Theorem 4.1, the continuous mapping theorem and $c_{1-\alpha}$ being a continuity point of the cdf of $\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|$. ■

Proof of Theorem 5.3: Since support functions are continuous in $p \in \mathbb{S}^{d_\theta}$, it follows that $\hat{\mathfrak{M}}_n(\theta) \subseteq \mathbb{S}^{d_\theta}$ is closed and bounded and therefore compact. Moreover, by Theorem 17.31 in Aliprantis and Border (2006), $\mathfrak{M}(\theta)$ is nonempty and compact valued, while Theorem 4.2 and Lemma A.9 in Kaido (2010) imply $d_H(\hat{\mathfrak{M}}_n(\theta), \mathfrak{M}(\theta)) = o_p(1)$. Therefore, Assumption 4.3 is satisfied with $\mathfrak{M}(\theta) = \Psi_0$ and $\hat{\mathfrak{M}}_n(\theta) = \hat{\Psi}_n$. In turn, by Theorem 11.1 in Davydov et al. (1998), the cdf of $\sup_{p \in \mathfrak{M}(\theta)} |\mathbb{G}_0(p)|_+$ is continuous and strictly increasing except possibly at zero. However, since $\mathfrak{M}(\theta)$ is nonempty, $\text{Var}\{\mathbb{G}_0(p)\} > 0$ for all $p \in \mathbb{S}^{d_\theta}$, (194) and $\alpha < 0.5$ implies that the cdf of $\sup_{p \in \mathfrak{M}(\theta)} |\mathbb{G}_0(p)|_+$ is continuous and strictly increasing at $c_{1-\alpha}(\theta)$. By Theorem 4.4 it then follows that $\hat{c}_{1-\alpha}(\theta) = c_{1-\alpha}(\theta) + o_p(1)$.

Suppose $\theta \in \Theta_0(P_0)^\circ$. Since $d_H(\Theta_0(P_0), \hat{\Theta}_n) = o_p(1)$ by Theorem 4.2, it follows that with probability tending to one $\theta \in \hat{\Theta}_n^\circ$. Therefore, $H_n(\theta) = 0$ with probability tending to one, and since $\hat{c}_{1-\alpha}(\theta) \xrightarrow{p} c_{1-\alpha}(\theta) > 0$, we conclude:

$$\liminf_{n \rightarrow \infty} P(\theta \in \hat{P}_n) = \liminf_{n \rightarrow \infty} P(H_n(\theta) \leq \hat{c}_{1-\alpha}(\theta)) = 1. \quad (198)$$

Suppose on the other hand that $\theta \in \partial\Theta_0(P_0)$. Theorem 4.1 and Lemma A.8 in Kaido (2010) then imply that:

$$H_n(\theta) \xrightarrow{L} \sup_{p \in \mathfrak{M}(\theta)} |-\mathbb{G}_0(p)|_+. \quad (199)$$

Therefore, since $\hat{c}_{1-\alpha}(\theta) \xrightarrow{p} c_{1-\alpha}(\theta)$ and the cdf of $\sup_{p \in \mathfrak{M}(\theta)} |\mathbb{G}_0(p)|$ is continuous at $c_{1-\alpha}(\theta)$ we obtain from (199):

$$\lim_{n \rightarrow \infty} P(\theta \in \hat{P}_n) = \lim_{n \rightarrow \infty} P(H_n(\theta) \leq \hat{c}_{1-\alpha}(\theta)) = P(\sup_{p \in \mathfrak{M}(\theta)} |-\mathbb{G}_0(p)|_+ \leq c_{1-\alpha}(\theta)) = 1 - \alpha, \quad (200)$$

which establishes the claim of the Theorem. ■

Proof of Theorem 5.4: We first establish the upper bound by showing $\{\pi_n\}$ corresponds to (local) α -level tests for:

$$H_0 : \{\nu(p_0, \{\theta_0\}) - \nu(p_0, \Theta_0(P_0))\} \leq 0 \quad H_1 : \{\nu(p_0, \{\theta_0\}) - \nu(p_0, \Theta_0(P_0))\} > 0. \quad (201)$$

Towards this end, let \mathbf{P} be as in Theorem B.1 and $\eta \mapsto h_\eta$ be a curve in L_μ^2 with $h_0 = \sqrt{dP_0/d\mu}$ satisfying $\int \tilde{l}(x) \dot{h}_0(x) h_0(x) d\mu(x) > 0$. Theorem 4.1 implies $\tilde{l}(x)$ is the efficient influence function for $\nu(p_0, \{\theta_0\}) - \nu(p_0, \Theta_0(P_0))$, and therefore, in order to establish $\{\pi_n\}$ corresponds to a (local) α -level test for (201), we need to show:

$$\limsup_{n \rightarrow \infty} \pi_n(h_{\eta/\sqrt{n}}) \leq \alpha \quad (202)$$

for any $\eta \leq 0$. Define the functional $\psi : \mathcal{C}(\mathbb{S}^{d_\theta}) \rightarrow \mathbf{R}$ to be pointwise given by $\psi(f) = \sup_{p \in \mathbb{S}^{d_\theta}} \{\nu(p, \{\theta_0\}) - f(p)\}$ and let P_η denote the measure satisfying $h_\eta = \sqrt{dP_\eta/d\mu}$. Since by Theorem 4.1 the estimator $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ is regular and asymptotically linear, Theorem 2.1 in van der Vaart (1991) implies $\eta \mapsto \nu(\cdot, \Theta_0(P_\eta))$ is pathwise differentiable.

Hence, by the chain rule, Theorem B.2 and Lemma A.7 in Kaido (2010) we obtain:

$$\frac{\partial}{\partial \eta} \psi(\nu(\cdot, \Theta_0(P_\eta))) \Big|_{\eta=0} = -2 \int \lambda(p_0, P_0)' H(\theta_0) m(x, \theta_0) \dot{h}_0(x) h_0(x) d\mu(x) > 0, \quad (203)$$

where the final result holds by definition of $\tilde{l}(x)$ and $\int \tilde{l}(x) \dot{h}_0(x) h_0(x) d\mu(x) > 0$. Therefore, since $\theta_0 \in \partial\Theta_0(P_0)$, $\psi(\nu(\cdot, \Theta_0(P_0))) = 0$ and hence (203) implies that $\text{sign}\{\psi(\nu(\cdot, \Theta_0(P_\eta)))\} = \text{sign}\{\eta\}$ for η small. We conclude that for η in a neighborhood of zero, $\theta_0 \in \Theta_0(P_\eta)$ if $\eta \leq 0$ and $\theta_0 \notin \Theta_0(P_\eta)$ if $\eta > 0$. Thus, $h_\eta \in \mathbf{H}(\theta_0)$ which implies (202) as a result of (61). By Theorem 18.12 in Kosorok (2008) we then finally obtain:

$$\limsup_{n \rightarrow \infty} \pi_n(h_{\eta/\sqrt{n}}) \leq 1 - \Phi(z_{1-\alpha} - \eta \frac{2E[\tilde{l}(X_i)\dot{h}_0(X_i)/h_0(X_i)]}{\sqrt{E[\mathbb{G}_0^2(p_0)]}}). \quad (204)$$

Next, we aim to show $\{\pi_n^*\}$ attains the bound. By Lemma A.7 in Kaido (2010), Theorem 4.1 and Theorem 18.6 in Kosorok (2008), $\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n)))\}$ is an efficient estimator for $\psi(\nu(\cdot, \Theta_0(P_0)))$ and hence it is regular. Let $L_{\eta/\sqrt{n}}$ denote the implied Law when $X_i \sim P_{\eta/\sqrt{n}}$ and note that the functional delta method and regularity then imply:

$$\sqrt{n}\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n))) - \psi(\nu(\cdot, \Theta_0(P_{\eta/\sqrt{n}})))\} \xrightarrow{L_{\eta/\sqrt{n}}} \mathbb{G}_0(p_0). \quad (205)$$

Therefore, since $\psi(\nu(\cdot, \Theta_0(P_0))) = 0$ due to $\theta_0 \in \partial\Theta_0(P_0)$, applying (203) and result (205) we then conclude:

$$\sqrt{n}\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n))) - \psi(\nu(\cdot, \Theta_0(P_0)))\} \xrightarrow{L_{h_{\eta/\sqrt{n}}}} \mathbb{G}_0(p_0) + \eta \int 2\tilde{l}(x)\dot{h}_0(x)h_0(x)d\mu(x). \quad (206)$$

As shown in the proof of Theorem 5.3, $\hat{c}_{1-\alpha}(\theta_0) = c_{1-\alpha}(\theta_0) + o_p(1)$ when $X_i \sim P_0$ and therefore by Theorem 11.14 in Kosorok (2008) also when $X_i \sim P_{\eta/\sqrt{n}}$. Hence, since $H_n(\theta) = \max\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n))), 0\}$ we obtain:

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{\eta/\sqrt{n}}(H_n(\theta_0) > \hat{c}_{1-\alpha}(\theta_0)) &= \lim_{n \rightarrow \infty} P_{\eta/\sqrt{n}}(\max\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n))), 0\} > c_{1-\alpha}(\theta_0)) \\ &= \lim_{n \rightarrow \infty} P_{\eta/\sqrt{n}}(\psi(\nu(\cdot, \Theta_0(\hat{P}_n))) > c_{1-\alpha}(\theta_0)) = P(\mathbb{G}_0(p_0) > c_{1-\alpha}(\theta_0) - 2\eta \int \tilde{l}(x)\dot{h}_0(x)h_0(x)d\mu(x)) \end{aligned} \quad (207)$$

where the second equality follows from $c_{1-\alpha}(\theta_0) > 0$ due to $\alpha < 0.5$ and the last equality is a result of (206); which verifies $\{\pi_n^*\}$ attains the bound in (62). Moreover, if $h_\eta \in \mathbf{H}(\theta_0)$, then (203) implies $\int \tilde{l}(x)\dot{h}_0(x)h_0(x)d\mu(x) > 0$ and hence from (207) we conclude that $\{\pi_n^*\}$ satisfies (61) when $\eta \leq 0$. ■

References

- AI, C. and CHEN, X. (2009). Semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. Working paper. Cowles Foundation for Economic Research - Yale University.
- ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis - A Hitchhiker's Guide*. Springer-Verlag, Berlin.
- ANDREWS, D. W. K. and GUGGENBERGER, P. (2009). Validity of subsampling and “plug-in asymptotic” inference for parameters defined by moment inequalities. *Econometric Theory*, **25** 669–709.
- ANDREWS, D. W. K. and JIA, P. (2008). Moment inequalities: A recommended moment selection procedure. Working paper. Cowles Foundation for Research in Economics - Yale University.
- ANDREWS, D. W. K. and SOARES, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, **78** 119–157.
- BARBE, P. and BERTAIL, P. (1995). *The Weighted Bootstrap*. Springer-Verlag, New York.

- BERESTEANU, A., MOLCHANOV, I. and MOLINARI, F. (2009). Sharp identification regions in models with convex predictions: Games, individual choice, and incomplete data. Working paper. Centre for Microdata Methods and Practice (cemmap).
- BERESTEANU, A. and MOLINARI, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, **76** 763–814.
- BHATIA, R. (1997). *Matrix Analysis*. Springer-Verlag, New York.
- BICKEL, P. J., KLASSEN, C. A., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.
- BONTEMPS, C., MAGNAC, T. and MAURIN, E. (2007). Set identified linear models. Working paper. Toulouse School of Economics.
- BROWN, B. W. and NEWEY, W. K. (1998). Efficient semiparametric estimation of expectations. *Econometrica*, **66** 453–464.
- BUGNI, F. A. (2010). Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica*, **78** 735–753.
- CANAY, I. A. (2010). El inference for partially identified models: Large deviations optimality and bootstrap validity. *Journal of Econometrics*, **156** 284–303.
- CHAMBERLAIN, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, **34** 305–334.
- CHAMBERLAIN, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica*, **60** 567–596.
- CHEN, X. and POUZO, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth moments. *Journal of Econometrics*, **152** 46–60.
- CHERNOZHUKOV, V., HONG, H. and TAMER, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, **75** 1243–1284.
- CHESHER, A. (2009). Single equation endogenous binary response models. Working paper. Centre for Microdata Methods and Practice (cemmap).
- CILIBERTO, F. and TAMER, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, **77** 1791–1828.
- DAVYDOV, Y. A., LIFSHITS, M. A. and SMORODINA, N. V. (1998). *Local Properties of Distributions of Stochastic Functionals*. American Mathematical Society, Providence.
- DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*. Springer, New York.
- HANSEN, L., HEATON, J. and LUTTMER, E. (1995). Econometric evaluation of asset pricing models. *Review of Financial Studies*, **8** 237–274.
- HIRANO, K. and PORTER, J. R. (2009). Impossibility results for nondifferentiable functionals. Working paper. University of Arizona.
- IMBENS, G. W. and MANSKI, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, **72** 1845–1857.

- KAIDO, H. (2010). A dual approach to inference for partially identified econometric models. Working Paper. Boston University.
- KLINE, P. and SANTOS, A. (2010). A score based approach to wild bootstrap inference. NBER Working Paper 16127.
- KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag, New York.
- LI, S., OGURA, Y. and KREINOVICH, V. (2002). *Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables (Theory and Decision Library B)*. Springer-Verlag, Berlin.
- LUENBERGER, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, New York.
- LUTTMER, E. (1996). Asset pricing in economies with frictions. *Econometrica*, **64** 1439–1467.
- MAHAJAN, A. (2003). Misclassified regressors in binary choice models. Working Paper. Stanford University.
- MANSKI, C. F. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag, New York.
- MANSKI, C. F. and PEPPER, J. (2000). Monotone instrumental variables. *Econometrica*, **68** 997–1010.
- MANSKI, C. F. and TAMER, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, **70** 519–546.
- MENZEL, K. (2008). Estimation and inference with many moment inequalities. Working paper. Massachusetts Institute of Technology.
- MENZEL, K. (2009). Consistent estimation with many moment inequalities. Working paper. New York University.
- MILGROM, P. and SEGAL, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, **70** 583–601.
- PAKES, A. (2010). Alternative models for moment inequalities. *Econometrica*, **78** 1783–1822.
- PAKES, A., PORTER, J., HO, K. and ISHII, J. (2006). Moment inequalities and their application. Working Paper, Harvard University.
- ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton.
- ROMANO, J. P. and SHAIKH, A. M. (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.
- ROMANO, J. P. and SHAIKH, A. M. (2010). Inference for the identified set in partially identified econometric models. *Econometrica*, **78** 169–211.
- ROSEN, A. M. (2009). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics*, **146** 107–117.
- SCHNEIDER, R. (1993). *Convex Bodies: the Brunn-Minkowski Theory (Encyclopedia of Mathematics and its Applications; Volume 44)*. Cambridge University Press, Cambridge.
- SONG, K. (2010). Robust estimation of some nonregular parameters. PIER Working Paper 10-20.
- STOYE, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica*, **77** 1299–1315.
- VAN DER VAART, A. (1991). On differentiable functionals. *The Annals of Statistics*, **19** 178–204.

VAN DER VAART, A. (1999). *Asymptotic Statistics*. Cambridge University Press, New York.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer-Verlag, New York.

Table 1: Proportion of simulated samples with expanded set estimators

		C			
		K	0.1	0.5	1
$n = 200$	5	0.000	0.000	0.000	
	10	0.290	0.000	0.000	
	15	0.782	0.010	0.000	
$n = 500$	5	0.000	0.000	0.000	
	10	0.051	0.000	0.000	
	15	0.410	0.000	0.000	
$n = 1,000$	5	0.000	0.000	0.000	
	10	0.003	0.000	0.000	
	15	0.149	0.000	0.000	

Table 2: Mean Hausdorff distance

		C			
		K	0.1	0.5	1
$n = 200$	5	0.153	0.150	0.151	
	10	0.275	0.250	0.250	
	15	0.514	0.361	0.359	
$n = 500$	5	0.094	0.094	0.094	
	10	0.177	0.155	0.155	
	15	0.360	0.218	0.219	
$n = 1,000$	5	0.066	0.066	0.066	
	10	0.130	0.109	0.109	
	15	0.201	0.154	0.154	

Table 3: Mean inner Hausdorff distance

		C		
	K	0.1	0.5	1
$n = 200$	5	0.137	0.144	0.144
	10	0.106	0.105	0.108
	15	0.308	0.045	0.047
$n = 500$	5	0.088	0.090	0.090
	10	0.085	0.093	0.094
	15	0.043	0.051	0.054
$n = 1,000$	5	0.063	0.064	0.064
	10	0.075	0.081	0.081
	15	0.061	0.055	0.053

Table 4: Mean outer Hausdorff distance

		C		
	K	0.1	0.5	1
$n = 200$	5	0.150	0.145	0.145
	10	0.273	0.249	0.250
	15	0.296	0.359	0.359
$n = 500$	5	0.092	0.090	0.090
	10	0.185	0.154	0.154
	15	0.360	0.218	0.219
$n = 1,000$	5	0.064	0.064	0.064
	10	0.130	0.108	0.108
	15	0.185	0.154	0.154

Table 5: Coverage probabilities of the one-sided confidence set

		C			
		K	0.1	0.5	1
$n = 200$	5	0.946	0.944	0.942	
	10	0.988	0.915	0.900	
	15	0.995	0.896	0.824	
$n = 500$	5	0.953	0.941	0.941	
	10	0.958	0.912	0.910	
	15	0.896	0.886	0.877	
$n = 1,000$	5	0.962	0.950	0.950	
	10	0.926	0.921	0.921	
	15	0.978	0.893	0.891	

Table 6: Coverage probabilities of the two-sided confidence set

		C			
		K	0.1	0.5	1
$n = 200$	5	0.958	0.945	0.943	
	10	0.996	0.953	0.936	
	15	0.909	0.948	0.886	
$n = 500$	5	0.964	0.942	0.942	
	10	0.983	0.953	0.951	
	15	0.949	0.938	0.931	
$n = 1,000$	5	0.971	0.949	0.949	
	10	0.963	0.952	0.952	
	15	0.987	0.892	0.939	

Figure 1: Identified Set as a Function of C and K

