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

First Draft: April, 2011
This Draft: December, 2012

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 wide class of asymptotic loss functions. In addition, we show the "plug-in" estimator is efficient, and devise a consistent bootstrap procedure for estimating its limiting distribution. The setting we examine is related to an incomplete linear model studied in Beresteanu and Molinari (2008) and Bontemps et al. (2012), which further enables us to establish the semiparametric efficiency of their proposed estimators for that problem.

Keywords: Semiparametric efficiency, partial identification, moment inequalities.

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

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., 2011), 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 vector with distribution P, $\Theta \subset \mathbf{R}^{d_\theta}$ denote the parameter space and $m : \mathcal{X} \times \Theta \to \mathbf{R}^{d_m}$ and $F : \mathbf{R}^{d_m} \to \mathbf{R}^{d_F}$ be known functions. In many models, the identified set is of the general form

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

A prevalent specification is one in which F is the identity mapping, in which case (1) reduces to the moment inequalities model studied in Chernozhukov 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).

We contribute to the existing literature by developing an asymptotic efficiency concept for estimating an important subset of these models. Heuristically, estimation of the identified set is tantamount to estimation of its boundary. In obtaining an asymptotic efficiency result, it is therefore instrumental to characterize the boundary of the identified set as a function of the unknown distribution P. We obtain such a characterization in the special, yet widely applicable, setting in which the constraint functions are convex, for example linear, in θ. In such instances, the identified set is itself convex and its boundary is determined by the hyperplanes that are tangent to it. The set of tangent, or supporting, hyperplanes can in turn be identified with a unique function on the unit sphere called the *support function* of the identified set. As a result, estimation of the identified set may be accomplished through the estimation of its support function – an insight previously exploited by Bontemps et al. (2012), 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. 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. its corresponding noise process Δ_0 equals zero almost surely. Obtaining a semiparametric efficiency bound then amounts to characterizing the distribution of $\mathbb{G}_0$, which in finite dimensional problems is equivalent to reporting its covariance matrix. In the present context, we obtain the semiparametric efficiency bound by deriving the covariance kernel of the Gaussian process $\mathbb{G}_0$. These insights are readily applicable to other convex partially identified models; a point we illustrate by showing estimators proposed in Beresteanu and Molinari (2008) and Bontemps et al. (2012) are efficient.

Among the implications of semiparametric efficiency, is that an efficient estimator minimizes diverse measures of asymptotic risk among regular estimators. Due to the close link between convex sets and their support functions, optimality in estimating the support function of the identified set further leads to optimality in estimating the identified set itself. Specifically, we show that, among regular convex set estimators, the set associated with the efficient estimator for the support function minimizes asymptotic risk for a wide class of loss functions based on Hausdorff distance. These results complement Song (2012), who employs semiparametric efficient estimators of interval valued identified sets to construct an optimal statistic for researchers that must make a point decision.

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

In related work, Beresteanu and Molinari (2008) first employ support functions in the study of partially identified models. The authors derive methods for conducting inference on the identified set through its support function, providing insights we rely upon in our analysis. The use of support functions to characterize semiparametric efficiency, however, is novel to this paper. Other work on estimation includes Hirano and Porter (2012), who 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 clarifying what the sources of irregularity are in setting where the parameter of interest has dimension greater than one.

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

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

2 General Setup

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

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

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

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

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

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

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

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

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

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

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

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

whenever $\theta_1^{(j)} \neq \theta_2^{(j)}$ for some $j \in \mathcal{S}_i$. For instance, if $\mathcal{S}_i = \emptyset$ then by (3), constraint i is linear in θ with known slope but intercept potentially depending on P. Similarly, if $\mathcal{S}_i = \{1,\ldots,d_\theta\}$, then

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

constraint i is strictly convex in θ. In between these cases are specifications of the constraints that are linear in some parameters and strictly convex in others.

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

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

2.1 Examples

In order to fix ideas, we briefly discuss applications of the general framework and refer the reader to the Supplemental Appendix for a more detailed analysis of these examples. For ease of exposition, we base our discussion 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

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

where $Z \in \mathbf{R}^{d_\theta}$ is a regressor with discrete support $\mathcal{Z} \equiv \{z_1, \ldots, z_K\}$ 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

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

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

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

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

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

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, the optimality conditions in (5) then imply θ_0 must satisfy the moment inequalities

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

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

Strictly convex moment inequalities arise in asset pricing (Hansen et al., 1995).[2]

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

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

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

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

Example 2.4 (Participation Constraint)**.** Consider an agent with Stone-Geary preferences over consumption $C \in \mathbf{R}_+$ and leisure $L \in [0, T]$ parametrized by

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

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

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

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

3 Semiparametric Efficiency

3.1 Preliminaries

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

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

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

[2] We 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.

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

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

Heuristically, the support function assigns to each vector p the signed distance between the origin and the hyperplane orthogonal to p that is tangent to K. By Hörmander's embedding theorem, the support functions of any two compact convex sets K_1 and K_2 belong to $\mathcal{C}(\mathbb{S}^{d_\theta})$ and in addition

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

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 isometry between them.

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

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

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

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

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

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

We let $\mathbf{M}$ denote the set of Borel probability measures on $\mathcal{X}$, endowed with the τ-topology,[3] and μ be a positive σ-finite measure such that P is absolutely continuous with respect to μ (denoted $P \ll \mu$). Of particular interest is the set $\mathbf{M}_\mu \equiv \{P \in \mathbf{M} : P \ll \mu\}$, which may be embedded in

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

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

$$\mathbf{S} \equiv \{h \in L^2_\mu : h = \sqrt{dQ/d\mu} \text{ for some } Q \in \mathbf{P}\} \ . \tag{10}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

where $\xrightarrow{L}$ denotes convergence in law, and Δ_0 is some tight random element independent of $\mathbb{G}_0$.[4]

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

[4] $\{T_n\}$ 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 $s \equiv \sqrt{dP/d\mu}$ 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 $P^n_{\eta_n}$.

bound is then equivalent to characterizing the distribution of $\mathbb{G}_0$. In finite dimensional problems, this amounts to computing the covariance matrix of the distributional limit. In the present context, we instead aim to obtain the covariance kernel for the Gaussian process $\mathbb{G}_0$, denoted

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

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

Remark 3.1. More generally, if a possibly nonconvex identified set $\Theta_0(P)$ is an element of a metric space $\mathbf{B}_1$, then we can consider estimation of the parameter $\rho_1 : \mathbf{P} \to \mathbf{B}_1$ given by $\rho_1(P) = \Theta_0(P)$. However, a key complication in this approach is that $\mathbf{B}_1$ is often not a vector space – a crucial requirement in the theory of semiparametric efficiency. For this reason, in our setting we instead employ an isometry $\rho_2 : \mathbf{B}_1 \to \mathbf{B}_2$ into a Banach space $\mathbf{B}_2$, and examine estimation of $\rho(P) \equiv \rho_2 \circ \rho_1(P)$.[5] This insight is applicable to other partially identified models – e.g. a bounded set K can be embedded in L^1_μ through its indicator function. Establishing pathwise weak differentiability in these contexts, however, will require substantially different arguments to ours. ∎

3.2 Efficiency Bound

3.2.1 Assumptions

We require the following assumptions to derive the distribution of the efficient limiting process $\mathbb{G}_0$.[6]

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

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

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

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

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

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

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

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

[6] We discuss the verification and implications of these assumptions for Examples 2.1-2.4 in the Appendix.

Assumption 3.6(i) implies $\Theta_0(P)$ is characterized by the inequality constraints and not by the parameter space. Certain parameter constraints, however, may be imposed through the moment restrictions; see Remark 3.4. In Assumption 3.6(ii), convexity of the constraints is required at all Q near P (in the τ-topology), which implies $\Theta_0(Q)$ is also convex. Assumptions 3.6(iii), together with Assumptions 3.4(ii) and 3.5(ii), ensure the constraints are differentiable in θ. Finally, Assumption 3.6(iv) is the key requirement ensuring $\nu(\cdot, \Theta_0(P))$ is a regular parameter at P. This assumption implies $\Theta_0(P)$ has a nonempty interior, which both rules out identification and enables us to obtain a Lagrangian representation for $\nu(\cdot, \Theta_0(P))$. Additionally, Assumption 3.6(iv) rules out moment equalities, though we note strictly convex moment equalities would imply the model is either identified or the identified set is nonconvex. Interestingly, a violation of Assumption 3.6(iv) is also the condition under which Hirano and Porter (2012) show irregularity in the problem they study.

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

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

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

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

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

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

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

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

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

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

$$\mathcal{T}_1(\theta, d, P) \equiv \sum_{i=1}^{d_F} \Big\{ d^{(i)} (F^{(i)}(\int m(x, \theta)dP(x)))^2 + (1 - d^{(i)})(F^{(i)}(\int m(x, \theta)dP(x)))^2_+ \Big\}$$

$$\mathcal{T}_2(\theta, \alpha, d, P) \equiv \sum_{j=1}^{d_\theta} \Big(\sum_{i=1}^{d_F} d^{(i)} \alpha^{(i)} \nabla F^{(i)}(\int m(x, \theta)dP(x)) \int \frac{\partial}{\partial \theta^{(j)}} m(x, \theta)dP(x) \Big)^2 ,$$

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

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

Remark 3.4. 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 $1 \leq i \leq d_F$ and all Q. Upper or lower bound constraints on individual elements $\theta^{(i)}$ of the vector θ may be similarly imposed. ■

3.2.2 Inverse Information Covariance Functional

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

$$\nu(p, \Theta_0(P)) = \sup_{\theta \in \Theta} \{\langle p, \theta\rangle \text{ s.t. } F(\int m(x,\theta)dP(x)) \leq 0\} \tag{11}$$

is the maximum of a convex program. Hence, $\nu(p, \Theta_0(P))$ also admits a Lagrangian representation

$$\nu(p, \Theta_0(P)) = \sup_{\theta \in \Theta} \{\langle p, \theta\rangle + \lambda(p, P)' F(\int m(x,\theta)dP(x))\} \ , \tag{12}$$

where, under our assumptions, the Lagrange multipliers are unique. Moreover, the maximizers of (11) also solve (12), and consist of the boundary points of $\Theta_0(P)$ at which $\Theta_0(P)$ is tangent to the hyperplane $\{\theta \in \mathbf{R}^{d_\theta} : \langle p, \theta\rangle = \nu(p, \Theta_0(P))\}$. These boundary points, together with their associated Lagrange multipliers, are instrumental in characterizing the semiparametric efficiency bound.

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

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

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

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

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

for any $\theta^*(\bar{p}) \in \arg\max_{\theta \in \Theta_0(P)} \langle \bar{p}, \theta\rangle$. Hence, since Lagrange multipliers corresponding to non-binding moment inequalities are zero, the semiparametric efficiency bound for $\nu(\bar{p}, \Theta_0(P))$ is the variance of a linear combination of the binding constraints at the boundary point $\theta^*(\bar{p}) \in \partial\Theta_0(P)$. Heuristically, the Lagrange multipliers represent the marginal value of relaxing the constraints in expanding the boundary of the identified set outwards in the direction $\bar{p}$ – 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 $\theta^*(\bar{p}) \in \partial\Theta_0(P)$.

3.3 Related Model

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

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

$$E[V(Y - Z'\theta_0)] = 0 \ , \tag{13}$$

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

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

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

$$\nu(p, \Theta_{0,I}(P)) = \int p'\Sigma(P)^{-1}v(y_L + 1\{p'\Sigma(P)^{-1}v > 0\}(y_U - y_L))dP(x) \ , \tag{14}$$

see Bontemps et al. (2012). We impose that Z and V be of equal dimension because it is only in this instance that (14) holds, which greatly simplifies verifying weak-pathwise differentiability.

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

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

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

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

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

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

$$\psi_\nu(p, x, P) \equiv \{y_L + 1\{p'\Sigma(P)^{-1}v > 0\}(y_U - y_L)\}v'\Sigma(P)^{-1}p \tag{15}$$

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

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

$$I^{-1}(p_1, p_2) = E[(\psi(p_1, X_i, P) - E[\psi(p_1, X_i, P)])(\psi(p_2, X_i, P) - E[\psi(p_2, X_i, P)])] \ .$$

[8] We thank an anonymous referee for this insight; see the Supplemental Appendix for a more detailed discussion.

The semiparametric efficiency bound of Theorem 3.3 coincides with the asymptotic distribution of the estimators studied in Beresteanu and Molinari (2008) and Bontemps et al. (2012), thus verifying their efficiency. We also note that if $P(Y_L = Y_U) = 1$, so that the model is identified, then Theorem 3.3 implies the efficient estimator is $p \mapsto \langle p, \hat{\theta} \rangle$ for $\hat{\theta}$ the GMM estimator of (13).

4 Efficient Estimation

4.1 The Estimator

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

$$\nu(p, \Theta_0(\hat{P}_n)) = \sup_{\theta \in \Theta} \ \{\langle p, \theta \rangle \text{ s.t. } F(\frac{1}{n}\sum_{i=1}^n m(X_i, \theta)) \leq 0\} \ . \tag{17}$$

It is useful to note Assumption 3.6(ii) implies the constraints in (17) are convex in $\theta \in \Theta$ with probability tending to one. As a result, $\nu(p, \Theta_0(\hat{P}_n))$ also admits a characterization as a Lagrangian

$$\nu(p, \Theta_0(\hat{P}_n)) = \sup_{\theta \in \Theta} \ \{\langle p, \theta \rangle + \lambda(p, \hat{P}_n)' F(\frac{1}{n}\sum_{i=1}^n m(X_i, \theta))\} \ . \tag{18}$$

This dual representation, together with the envelope theorem of Milgrom and Segal (2002), enables us to conduct a Taylor expansion of $\nu(\cdot, \Theta_0(\hat{P}_n))$ around $\nu(\cdot, \Theta_0(P))$. In this manner, we are able to characterize the influence function of $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ (in $\mathcal{C}(\mathbb{S}^{d_\theta})$), and establish its efficiency.

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

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

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

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

where $\mathbb{G}_0$ is a mean zero tight 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 for the limiting distribution of statistics $\{T_n(\theta)\}$ to be discontinuous in $\theta \in \Theta_0(P)$. It is interesting to note that, in contrast, in Theorem 4.1 $\mathbb{G}_0$ is continuous in $p \in \mathbb{S}^{d_\theta}$ almost surely.[9] Heuristically, the continuity of $\mathbb{G}_0$ results from the Lagrange multipliers determining the weight a binding constraint receives at each $p \in \mathbb{S}^{d_\theta}$. Hence, if p_1 and p_2 are close, then the complementary slackness condition and continuity of $p \mapsto \lambda(p, P)$ imply that

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

constraints that are binding at p_1 but not p_2 must have a correspondingly small weight. As a result, the empirical process is continuous despite different constraints being binding at different $p \in \mathbb{S}^{d_\theta}$.

4.2 Asymptotic Risk

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

Theorem 4.2. *Let Assumption 3.1-3.5 hold, $P \in \mathbf{P}$ and $L : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}_+$ be a subconvex function*[10] *such that for all $f \in \mathcal{C}(\mathbb{S}^{d_\theta})$, $L(f) \leq M_0 + M_1 \|f\|_\infty^\kappa$ for some $M_0, M_1 > 0$ and $\kappa < \infty$. If $\mathbb{D}_0$ are the continuity points of L and $P(\mathbb{G}_0 \in \mathbb{D}_0) = 1$, then for any regular estimator $\{T_n\}$ of $\nu(\cdot, \Theta_0(P))$:*

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

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

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

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

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

$$\hat{\Theta}_n \equiv co\{\Theta_0(\hat{P}_n)\} , \tag{19}$$

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

Corollary 4.1. *Let Assumption 3.1-3.5 hold, $P \in \mathbf{P}$ and $L : \mathbf{R}_+ \to \mathbf{R}_+$ be a subconvex function continuous on $\mathbb{D}_0 \subseteq \mathbf{R}_+$, and satisfying* $\limsup_{a\to\infty} L(a)a^{-\kappa} < \infty$ *for some $\kappa > 0$. If $\{K_n\}$ is a*

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

regular convex compact valued set estimator for $\Theta_0(P)$*, and* $P(\|\mathbb{G}_0\|_\infty \in \mathbb{D}_0) = 1$*, then*[11]

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

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

4.3 Marginal Identified Sets

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

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

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

Heuristically, Assumption 4.1 imposes that at each $p \in \mathbb{C}$ the corresponding tangent hyperplane be supported by a unique boundary point of $\Theta_0(P)$. In Remark 3.2, for instance, $\nu(p, \Theta_0(P))$ is pathwise weak differentiable, except precisely at the point $p \in \mathbb{S}^{d_\theta}$ for which the tangent hyperplane coincides with a "flat face" of $\Theta_0(P)$. To reflect this additional restriction on P, we define

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

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

$$m(x, \theta) \equiv (m_S(x, \theta)', m_A(x)', \theta')' \ . \tag{20}$$

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

$$F(\int m(x, \theta) dP(x)) = F_A(\int m_A(x) dP(x))\theta + F_S(\int m_S(x, \theta) dP(x)) \tag{21}$$

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

Assumption 4.2. *(i) The functions* $m : \mathcal{X} \times \Theta \to \mathbf{R}^{d_m}$ *and* $F : \mathbf{R}^{d_m} \to \mathbf{R}^{d_F}$ *satisfy* (20) *and* (21)*; (ii) For each* $i \in \{1, \ldots, d_F\}$*, we have either* $\mathcal{S}_i = \emptyset$ *or* $\mathcal{S}_i = \{1, \ldots, d_\theta\}$.

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

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

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

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

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

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

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

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

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

5 A Consistent Bootstrap

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

$$G_n^*(p) \equiv \lambda(p, \hat{P}_n)' \nabla F(\frac{1}{n}\sum_{i=1}^n m(X_i, \hat{\theta}(p))) \frac{1}{\sqrt{n}} \sum_{i=1}^n \{m(X_i, \hat{\theta}(p)) - \frac{1}{n}\sum_{i=1}^n m(X_i, \hat{\theta}(p))\} W_i \ , \tag{22}$$

where $\lambda(p, \hat{P}_n)$ is as in (18) and $\hat{\theta}(p)$ is any maximizer for the optimization problem in (18). Heuristically, the stochastic process $p \mapsto G_n^*(p)$ is constructed by perturbing an estimate of the efficient influence function (or score) by the random weights $\{W_i\}_{i=1}^n$. These weights are assumed to satisfy

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

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

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

5.1 Estimating Critical Values

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

$$\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p)) \ , \tag{23}$$

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

Assumption 5.2. *(i) $\Upsilon : \mathbf{R} \to \mathbf{R}$ is continuous; (ii) $\{\hat{\Psi}_n\}$ 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 (23) 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 (18). ■

Step 2: Generate a random sample $\{W_i\}_{i=1}^n$ satisfying Assumption 5.1 to construct G_n^*. ■

Step 3: Employing G_n^* and $\{\hat{\Psi}_n\}$, estimate the $1-\alpha$ quantile of $\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p))$ by

$$\hat{c}_{1-\alpha} \equiv \inf\{c : P(\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) \leq c \ |\{X_i\}_{i=1}^n) \geq 1-\alpha\} \ . \tag{24}$$

In practice, $\hat{c}_{1-\alpha}$ is often not explicitly computable but obtainable through simulation. ■

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

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

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

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

$$\lim_{n \to \infty} P(\Theta_0(P) \subseteq \hat{\Theta}_n^{\hat{c}_{1-\alpha}/\sqrt{n}}) = 1-\alpha \tag{25}$$

for any consistent estimator $\hat{c}_{1-\alpha}$ for $c_{1-\alpha}$. In particular, by letting $\Upsilon(a) = (-a)_+$, and $\hat{\Psi}_n = \Psi_0 = \mathbb{S}^{d_\theta}$, Theorem 5.2 implies (25) holds if $c_{1-\alpha}$ is estimated employing the proposed bootstrap. Alternatively, Chernozhukov et al. (2012) provide a related construction based on the efficient estimator that is equivariant to transformations of the parameters. ■

5.2 Application to Testing

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

$$H_0 : \theta \in \Theta_0(P) \qquad H_1 : \theta \notin \Theta_0(P) \ , \tag{26}$$

which is commonly inverted to construct confidence regions that covers each element of $\Theta_0(P)$ with a prespecified probability. In a related setting, Kaido (2010) tests (26) employing the statistic[13]

$$J_n(\theta) \equiv \sqrt{n}\vec{d}_H(\{\theta\}, \hat{\Theta}_n) \ . \tag{27}$$

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

$$c_{1-\alpha}(\theta) \equiv \inf\{c : P(\sup_{p \in \mathfrak{M}(\theta)} (-\mathbb{G}_0(p))_+ \leq c) \geq 1 - \alpha\} \ . \tag{28}$$

Estimating $c_{1-\alpha}(\theta)$ requires a consistent estimator for $\mathfrak{M}(\theta)$, for which Kaido (2010) proposes

$$\hat{\mathfrak{M}}_n(\theta) \equiv \{p \in \mathbb{S}^{d_\theta} : \{\nu(p, \{\theta\}) - \nu(p, \Theta_0(\hat{P}_n))\} \geq \sup_{\tilde{p} \in \mathbb{S}^{d_\theta}} \{\nu(\tilde{p}, \{\theta\}) - \nu(\tilde{p}, \Theta_0(\hat{P}_n))\} - \frac{\kappa_n}{\sqrt{n}}\} \ , \tag{29}$$

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 5.2 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

$$\hat{c}_{1-\alpha}(\theta) \equiv \inf\{c : P(\sup_{p \in \hat{\mathfrak{M}}_n(\theta)} (-G_n^*(p))_+ \leq c \,|\{X_i\}_{i=1}^n) \geq 1 - \alpha\} \ . \tag{30}$$

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

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

$$\liminf_{n \to \infty} P(J_n(\theta) \leq \hat{c}_{1-\alpha}(\theta)) \geq 1 - \alpha \ . \tag{31}$$

5.2.1 Local Properties

The test that rejects (26) whenever $J_n(\theta) > \hat{c}_{1-\alpha}(\theta)$ satisfies a local optimality property. Specifically, we show the power function of any test that controls size over local parametric submodels must be weakly smaller than that of a test based on $J_n(\theta)$ for all $\theta \in \partial\Theta_0(P)$ that are supported by a unique hyperplane. Formally, let $h_\eta = \sqrt{dP_\eta/d\mu}$ and $\mathbf{H}(\theta)$ denote the set of submodels $\eta \mapsto P_\eta$ in $\mathbf{P}$ with

$$\text{(i) } h_0 = \sqrt{dP/d\mu} \qquad \text{(ii) } \theta \in \Theta_0(P_\eta) \text{ if } \eta \leq 0 \qquad \text{(iii) } \theta \notin \Theta_0(P_\eta) \text{ if } \eta > 0 \ . \tag{32}$$

Thus, $\mathbf{H}(\theta)$ is the set of submodels passing through P for which P_η satisfies the null hypothesis in (26) for $\eta \leq 0$, and the alternative for $\eta > 0$. We consider tests in terms of their power functions $\pi : \mathbf{H}(\theta) \to [0, 1]$, where $\pi(P_\eta)$ is the probability the null hypothesis is rejected when $X_i \sim P_\eta$.

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

$$\limsup_{n \to \infty} \pi_n(P_{\eta/\sqrt{n}}) \leq \alpha \ . \tag{33}$$

[13]Kaido (2010) examines an arbitrary estimator of $\nu(\cdot, \Theta_0(P))$, not necessarily the efficient one. This type of test statistic was first studied by Bontemps et al. (2012) in the context of the incomplete linear model of Section 3.3.

If $\{\pi_n^*\}$ *is the power function of the test that rejects when* $J_n(\theta_0) > \hat{c}_{1-\alpha}(\theta_0)$*, then* $\{\pi_n^*\}$ *satisfies* (33)*. Moreover, for* $\tilde{l}(x) \equiv -\lambda(p_0, P)' H(\theta_0)\{m(x, \theta_0) - E[m(X_i, \theta_0)]\}$ *and any* $P_\eta \in \mathbf{H}(\theta_0)$*,* $\eta > 0$

$$\limsup_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}}) \leq \lim_{n\to\infty} \pi_n^*(P_{\eta/\sqrt{n}}) = 1 - \Phi\Big(z_{1-\alpha} - \eta \frac{2E[\tilde{l}(X_i)\dot{h}_0(X_i)/h_0(X_i)]}{\sqrt{E[\mathbb{G}_0^2(p_0)]}}\Big) , \tag{34}$$

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

The null hypothesis in (26) holds if and only if $\langle p, \theta\rangle \leq \nu(p, \Theta_0(P))$ for all $p \in \mathbb{S}^{d_\theta}$. When $\mathfrak{M}(\theta) = \{p_0\}$, such inequality holds with equality only at p_0. Heuristically, any local perturbation $P_{\eta/\sqrt{n}}$ of P that violates the null hypothesis in (26) must then satisfy $\langle p_0, \theta\rangle > \nu(p_0, \Theta_0(P_{\eta/\sqrt{n}}))$. As a result, it is possible to locally relate (26) to the problem of testing $\langle p_0, \theta\rangle \leq \nu(p_0, \Theta_0(P))$ against $\langle p_0, \theta\rangle > \nu(p_0, \Theta_0(P))$. The limiting experiment of the latter hypothesis is akin to a one sided test for a mean, and Theorem 5.4 follows by showing the proposed test is optimal in this context. We note, however, that the size control requirement in (33) is local to a $P \in \mathbf{P}$, and the proposed test does not necessarily control size uniformly over a larger set of distributions.

6 Conclusion

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

References

Ai, C. and Chen, X. (2009). Semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. Working paper. Yale University.

Andrews, D. W. K. and Barwick, P. J. (2012). Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica*, **80** 2805–2826.

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 Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, **78** 119–157.

Beresteanu, A., Molchanov, I. and Molinari, F. (2011). Sharp identification regions in models with convex predictions: Games, individual choice, and incomplete data. *Econometrica*, **79** 1785–1821.

Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, **76** 763–814.

Bickel, P. J., Klassen, C. A., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York.

Blundell, R. and Macurdy, T. (1999). Labor supply: A review of alternative approaches. In *Handbook of Labor Economics 3A* (O. Ashenfelter and D. Card, eds.). North Holland, Elsevier.

Bontemps, C., Magnac, T. and Maurin, E. (2012). Set identified linear models. *Econometrica*, **80** 1129–1155.
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.
Bugni, F. A., Canay, I. A. and Shi, X. (2012). Specification test in partially identified models defined by moment inequailities. Working paper. Northwestern University.
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.
Chernozhukov, V., Hong, H. and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, **75** 1243–1284.
Chernozhukov, V., Kocatulum, E. and Menzel, K. (2012). Inference on sets in finance. Working paper. Massachusetts Institute of Technology.
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.
Donald, S. G. and Hsu, Y. (2009). Estimation and inference for distriubtion functions and quantile functions in treatment effect models. Working paper. University of Texas - Austin.
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. (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, **80** 1769–1790.
Kaido, H. (2010). A dual approach to inference for partially identified econometric models. Working Paper. Boston University.
Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, **1** 23–41.
Lewbel, A. (1995). Consistent nonparametric hypothesis tests with an application to slutsky symmetry. *Journal of Econometrics*, **67** 379–401.
Luttmer, E. (1996). Asset pricing in economies with frictions. *Econometrica*, **64** 1439–1467.
Mahajan, A. (2003). Misclassified regressors in binary choice models. Working Paper. UCLA.
Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer, New York.
Manski, C. F. and Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, **70** 519–546.
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.
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.
Song, K. (2010). Robust estimation of some nonregular parameters. PIER Working Paper 10-20.
Song, K. (2012). Point decisions for interal-identified parameters. Working Paper. UBC.
van der Vaart, A. (1991). On differentiable functionals. *The Annals of Statistics*, **19** 178–204.

Supplemental Appendix

In this Supplemental Appendix we include all proofs of results stated in the main text, a more detailed discussion of the examples introduced in Section 2.1, and the results of our Monte Carlo study. The proof of each main result is contained in its own Appendix, which also includes a discussion of the strategy of proof and the role of the auxiliary results. The contents of the Supplemental Appendix are organized as follows:

Appendix A: Contains the proof of Theorem 3.2 and required auxiliary results.

Appendix B: Contains the proofs of Theorems 4.1, 4.2, Corollary 4.1 and required auxiliary results.

Appendix C: Contains the proof of Theorem 4.3 and required auxiliary results.

Appendix D: Contains proof of Theorems 5.1, 5.2, 5.3 and 5.4.

Appendix E: Contains the proof of Theorem 3.3, and a discussion of regularity in the incomplete linear model.

Appendix F: Discusses our Assumptions in the context of Examples 2.1, 2.2, 2.3 and 2.4.

Appendix G: Reports the results of the Monte Carlo study.

For ease of reference, the following list includes notation and definitions that will be used in 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 $\mathcal{X} \subseteq \mathbf{R}^{d_X}$.
$\mathbf{M}_\mu$	For some $\mu \in \mathbf{M}$, the set $\mathbf{M}_\mu \equiv \{P \in \mathbf{M} : P \ll \mu\}$.
$N(Q)$	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 $\|\cdot\|$.
$\mathcal{S}_i$	The coordinates of θ on which $\theta \mapsto F_S^{(i)}(\int m_S(x,\theta)dP(x))$ depends.
$\Xi(p, Q)$	The set of maximizers of $\sup_{\theta \in \Theta} \langle p, \theta \rangle$ s.t. $F(\int m(x,\theta)dQ(x)) \leq 0$.

Appendix A - Proof of Theorem 3.2

This Appendix contains the proof of Theorem 3.2. Several of the auxiliary results are stated in more generality than needed so that they may be employed in the derivations in Theorems 4.1 and 4.3 as well.

The proof of Theorem 3.2 proceeds by verifying the conditions of Theorem 5.2.1 in Bickel et al. (1993), which requires two key ingredients: (i) Characterizing the tangent space at P, which we accomplish in Theorem A.1, and (ii) Showing $Q \mapsto \nu(\cdot, \Theta_0(Q))$ is weak-pathwise differentiable at P, which we verify in Theorem A.2. Before proceeding to the formal derivation of these results, we provide an outline of the general structure of the proof.

Tangent Space (Theorem A.1)

Step 1: Lemma A.16 establishes that if $\mathbf{P}$ is open relative to $\mathbf{M}_\mu$ in the τ-topology, then the tangent space must be unrestricted. Intuitively, if $\mathbf{P}$ is open and $P \in \mathbf{P}$, then all distribution Q close to P must also be in $\mathbf{P}$. Therefore knowing that $P \in \mathbf{P}$ does not contain information that may be exploited in estimation. ■

Step 2: Theorem A.1 then follows from establishing that there is a neighborhood $N(P)$ of P such that all $Q \in N(P)$ satisfy: (i) Assumption 3.6(i) (shown in Corollary A.3), (ii) Assumption 3.6(ii) (by hypothesis), (iii) Assumption 3.6(iii) (established in Lemma A.2), and (iv) Assumption 3.6(iv) (demonstrated in Lemma A.8). ■

Differentiability (Theorem A.2)

Step 1: Exploiting Lemma A.3, Lemma A.4 first shows $\Theta_0(P)$ has nonempty interior. Corollary A.2 then extends this result to hold for all Q in a neighborhood of $N(P)$ of P. ■

Step 2: Next, we note that since $\Theta_0(Q)$ has non empty interior for all $Q \in N(P)$, the support function has a saddle point representation. This is shown in Lemma A.9 which also establishes the Lagrange multipliers are unique. ■

Step 3: Lemma A.14 then employs the saddle point representation, the envelope theorem and auxiliary Lemma A.10, to show $P \mapsto \nu(p, \Theta_0(P))$ is pathwise weak-differentiable at any $p \in \mathbb{S}^{d_\theta}$. ■

Step 4: Finally, Theorem A.2 is shown by extending the pointwise result of Lemma A.14. The arguments exploit the continuity of Lagrange multipliers (Lemma A.12), and an auxiliary measurability result (Lemma A.13). ■

Lemma A.1. *Let $f : \mathcal{X} \times \Theta \to \mathbf{R}$ be a measurable function, bounded in $(x, \theta) \in \mathcal{X} \times \Theta$ and such that $\theta \mapsto f(x, \theta)$ is equicontinuous in $x \in \mathcal{X}$. If Assumption 3.2 holds and $\{Q_\alpha\}_{\alpha \in \mathfrak{A}}$ is a net in $\mathbf{M}$ with $Q_\alpha \to Q$, then:*

$$\limsup_{\alpha} \sup_{\theta \in \Theta} \left| \int f(x, \theta) dQ_\alpha(x) - \int f(x, \theta) dQ(x) \right| = 0 \ .$$

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:

$$\sup_{x \in \mathcal{X}, \tilde{\theta} \in N_{\delta(\theta)}(\theta)} |f(x, \theta) - f(x, \tilde{\theta})| < \epsilon \ . \tag{A.1}$$

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

$$\left| \int f(x, \theta) dQ_\alpha(x) - \int f(x, \theta) dQ(x) \right| \leq 2\epsilon + \max_{1 \leq i \leq K} \left| \int f(x, \theta_i)(dQ_\alpha(x) - dQ(x)) \right| \tag{A.2}$$

for any $\theta \in \Theta$. Since ϵ is arbitrary and $\max_{1 \leq i \leq K} |\int f(x, \theta_i)(dQ_\alpha(x) - dQ(x))| \to 0$ due to f being measurable and bounded for all θ, and $Q_\alpha \to Q$ in the τ-topology, the claim of the Lemma then follows from (A.2). ■

Lemma A.2. *If Assumptions 3.2, 3.4(i)-(ii) and 3.5 hold, then it follows that for every $P \in \mathbf{P}$ there is a neighborhood $N(P) \subseteq \mathbf{M}$ such that for all $Q \in N(P)$: $\{\int m(x, \theta) dQ(x)\}_{\theta \in \Theta}$ is compact and $\{\int m(x, \theta) dQ(x)\}_{\theta \in \Theta} \subset V_0$.*

Proof: First note Assumptions 3.4(i)-(ii) and the dominated convergence theorem imply that for any $Q \in \mathbf{M}$:

$$\lim_{\theta_1 \to \theta_2} \int m(x, \theta_1) dQ(x) = \int m(x, \theta_2) dQ(x) \ . \tag{A.3}$$

Thus, since Θ is closed by Assumption 3.2(i), result (A.3) implies the set $\mathcal{R}(Q) \equiv \{\int m(x, \theta) dQ(x)\}_{\theta \in \Theta}$ is closed in $\mathbf{R}^{d_m}$. Moreover, $\mathcal{R}(Q)$ is also bounded by Assumption 3.4(i), and hence we conclude $\mathcal{R}(Q)$ is compact, which establishes the first claim of the Lemma. Defining $\mathcal{R}(P)^\delta \equiv \{v \in \mathbf{R}^{d_m} : \inf_{\tilde{v} \in \mathcal{R}(P)} \|v - \tilde{v}\| < \delta\}$, it then follows from V_0 being open by Assumption 3.5, $\mathcal{R}(P)$ being compact, and Assumption 3.6(iii) that $\mathcal{R}(P) \subset V_0$. Hence, there exists a $\delta_0 > 0$ such that $\mathcal{R}(P)^{\delta_0} \subset V_0$, and the second claim of the Lemma then follows from Lemma A.1 implying there exists a $N(P)$ such that $\mathcal{R}(Q) \subseteq \mathcal{R}(P)^{\delta_0}$ for all $Q \in N(P)$. ■

Corollary A.1. *Let Assumptions 3.2, 3.4, 3.5 hold and $P \in \mathbf{P}$. Then there exists a neighborhood $N(P) \subseteq \mathbf{M}$ such that $F(\int m(x, \cdot) dQ(x)) : \Theta \to \mathbf{R}^{d_F}$ is continuously differentiable for all $Q \in N(P)$, and in addition:*

$$\nabla_\theta \{F(\int m(x, \theta) dQ(x))\} = \nabla F(\int m(x, \theta) dQ(x)) \int \nabla_\theta m(x, \theta) dQ(x) \ .$$

Proof: By Lemma A.2, there is a neighborhood $N(P) \subseteq \mathbf{M}$ such that $\int m(x, \theta) dQ(x) \in V_0$ for all $(\theta, Q) \in \Theta \times N(P)$. For any $Q \in N(P)$ and any $1 \leq i \leq d_F$, Assumption 3.5 then allows us to conclude that:

$$\nabla_\theta \{F^{(i)}(\int m(x, \theta) dQ(x))\} = \nabla F^{(i)}(\int m(x, \theta) dQ(x)) \int \nabla_\theta m(x, \theta) dQ(x) \ , \tag{A.4}$$

where the exchange of order of integration and differentiation is warranted by the mean value theorem, the dominated convergence theorem and Assumption 3.4(ii). Moreover, by Assumptions 3.4(i)-(iii) and 3.5(ii) we have:

$$\lim_{\theta_n \to \theta_0} \nabla F^{(i)}(\int m(x,\theta_n)dQ(x)) \int \nabla_\theta m(x,\theta_n)dQ(x) = \nabla F^{(i)}(\int m(x,\theta_0)dQ(x)) \int \nabla_\theta m(x,\theta_0)dQ(x) \quad \text{(A.5)}$$

by the dominated convergence theorem for any $\theta_n, \theta_0 \in \Theta$. The Corollary then follows from (A.4) and (A.5). ■

Lemma A.3. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold and* $P \in \mathbf{P}$. *It then follows that for every* $j \in \{1, \ldots, d_\theta\}$ *and every* $\theta_0 \in \Theta_0(P)$ *there exists a* $\theta_A \in \Theta_0(P)$ *satisfying* $\theta_0^{(j)} \neq \theta_A^{(j)}$.

Proof: The proof is by contradiction. Suppose $\theta_0 \in \Theta_0(P)$ and that for some $\bar{j} \in \{1, \ldots, d_\theta\}$ we have $\theta^{(\bar{j})} = \theta_0^{(\bar{j})}$ for all $\theta \in \Theta_0(P)$. Further define $K_i \equiv \{\theta \in \Theta : F^{(i)}(\int m(x,\theta)dP(x)) \leq 0\}$ and for any $A \subseteq \Theta$ let:

$$\Pi_{\bar{j}}\{A\} \equiv \{c \in \mathbf{R} : c = \theta^{(\bar{j})} \text{ for some } \theta \in A\} \ . \quad \text{(A.6)}$$

Since Θ is convex and $F^{(i)}(\int m(x,\cdot)dP(x)) : \Theta \to \mathbf{R}$ is convex by Assumptions 4.2(i), 3.6(ii) and $P \in \mathbf{P}$, it follows that K_i and $\bigcap_{i \in \mathcal{A}(\theta_0,P)} K_i$ are convex. Thus, $F^{(i)}(\int m(x,\theta_0)dP(x)) < 0$ for all $i \in \{1, \ldots, d_F\} \setminus \mathcal{A}(\theta_0, P)$ implies

$$\{\theta_0^{(\bar{j})}\} = \Pi_{\bar{j}}\{\bigcap_{i \in \mathcal{A}(\theta_0,P)} K_i\} \ , \quad \text{(A.7)}$$

or otherwise there would be a $\theta_A \in \Theta_0(P)$ with $\theta_A^{(\bar{j})} \neq \theta_0^{(\bar{j})}$. Moreover, Corollary A.1 and $P \in \mathbf{P}$ satisfying Assumption 3.6(iv) imply $\nabla_\theta\{F^{(i)}(\int m(x,\theta_0)dP(x))\} \neq 0$ for all $i \in \mathcal{A}(\theta_0, P)$. Hence, for each $i \in \mathcal{A}(\theta_0, P)$ there is a $\theta_i \in \Theta$ with

$$F^{(i)}(\int m(x,\theta_i)dP(x)) < 0 \quad \text{(A.8)}$$

due to $\theta_0 \in \Theta^o$ by $P \in \mathbf{P}$ satisfying Assumption 3.6(i). Let $\iota : \mathcal{A}(\theta_0, P) \to \{1, \ldots, \#\mathcal{A}(\theta_0, P)\}$ be a bijection and:

$$k^* \equiv \inf_{1 \leq k \leq \#\mathcal{A}(\theta_0,P)} k : \{\Pi_{\bar{j}}\{\bigcap_{i:\iota(i) \leq k} K_i\} = \{\theta_0^{(\bar{j})}\}\} \ , \quad \text{(A.9)}$$

where we note $2 \leq k^* \leq \#\mathcal{A}(\theta_0, P)$ due to (A.7) and $\{\Pi_{\bar{j}}\{K_i\}\}^o \neq \emptyset$ for all $i \in \mathcal{A}(\theta_0, P)$ by (A.8). Next, define:

$$\bar{K} \equiv \bigcap_{i:\iota(i) \leq k^*-1} K_i \qquad K_{i^*} \equiv K_{\iota^{-1}(k^*)} \ . \quad \text{(A.10)}$$

Since $\Pi_{\bar{j}}\{\bar{K}\}$ is not singleton valued, there exists a $\theta_A \in \bar{K}$ with $\theta_A^{(\bar{j})} \neq \theta_0^{(\bar{j})}$. It follows that if $\bar{\theta} \in \bar{K} \cap K_{i^*}$ then $\bar{\theta} \notin K_{i^*}^o$ for otherwise $c\theta_A + (1-c)\bar{\theta} \in \bar{K} \cap K_{i^*}$ for $c \in (0,1)$ sufficiently small, contradicting (A.9). We therefore conclude that $\bar{K} \cap K_{i^*}^o = \emptyset$, and by Theorem 5.12.3 in Luenberger (1969) that there is a $p^* \in \mathbb{S}^{d_\theta}$ such that:

$$\sup_{\theta \in K_{i^*}} \langle \theta, p^* \rangle \leq \inf_{\theta \in \bar{K}} \langle \theta, p^* \rangle \ . \quad \text{(A.11)}$$

Further note that both the infimum and supremum in (A.11) are attained at θ_0, and that since $P \in \mathbf{P}$ must satisfy Assumptions 3.6(iv), that $\{\nabla_\theta\{F^{(i)}(\int m(x,\theta_0)dP(x))\}\}_{i \in \mathcal{A}(\theta_0,P)}$ are linearly independent by Corollary A.1. Thus, it follows from Theorem 9.4.1 in Luenberger (1969) and $\theta_0 \in \Theta^o$ by $P \in \mathbf{P}$ satisfying Assumption 3.6(i) that:

$$\begin{aligned} 0 &= p^* + \gamma_0 \nabla_\theta\{F^{(\iota^{-1}(k^*))}(\int m(x,\theta_0)dP(x))\} \\ 0 &= p^* + \sum_{k=1}^{k^*-1} \gamma_k \nabla_\theta\{F^{(\iota^{-1}(k))}(\int m(x,\theta_0)dP(x))\} \end{aligned} \quad \text{(A.12)}$$

for some scalar $\gamma_0 \neq 0$ and vector $(\gamma_1, \ldots, \gamma_{k^*-1}) \neq 0$. However, result (A.12) and Corollary A.1 contradict $P \in \mathbf{P}$ satisfying Assumption 3.6(iv) and hence the Lemma follows. ■

Lemma A.4. *If Assumptions 3.2, 3.4, 3.5, 4.2(i) hold and* $P \in \mathbf{P}$, *then there exists a* $\theta_0 \in \Theta$ *such that:*

$$F^{(i)}(\int m(x,\theta_0)dP(x)) < 0 \quad \textit{for all } 1 \leq i \leq d_F \ .$$

Proof: Let $2^{\{1,\ldots,d_F\}}$ denote the power set of $\{1,\ldots,d_F\}$ and note that $\mathcal{A}(\cdot,P):\Theta\to 2^{\{1,\ldots,d_F\}}$. Since $\mathcal{A}(\cdot,P)$ has finite range, there exists a collection $\{\theta_j\}_{j=1}^J$ with $J<\infty$ and $\theta_j\in\Theta_0(P)$ such that for all $\theta\in\Theta_0(P)$:

$$\mathcal{A}(\theta,P)\in\{\mathcal{A}(\theta_j,P)\}_{j=1}^J\ . \tag{A.13}$$

Next, select weights $\{w_j\}_{j=1}^J$ such that $w_j>0$ and $\sum_j w_j=1$, and define $\theta_0\equiv\sum_j w_j\theta_j$. By convexity we obtain

$$F^{(i)}(\int m(x,\theta_0)dP(x))\le\sum_{j=1}^J w_jF^{(i)}(\int m(x,\theta_j)dP(x)) \tag{A.14}$$

for any $1\le i\le d_F$, which implies $\theta_0\in\Theta_0(P)$. Moreover, since $w_j>0$ for all $1\le j\le J$, it also follows that $F^{(i)}(\int m(x,\theta_0)dP(x))=0$ if and only if $F^{(i)}(\int m(x,\theta_j)dP(x))=0$ for all $1\le j\le J$. Thus, by (A.13) we conclude:

$$\mathcal{A}(\theta_0,P)=\bigcap_{j=1}^J\mathcal{A}(\theta_j,P)=\bigcap_{\theta\in\Theta_0(P)}\mathcal{A}(\theta,P)\ . \tag{A.15}$$

Next we aim to show $\mathcal{A}(\theta_0,P)=\emptyset$ which yields the claim of the Lemma. Toward this end, note that for any $1\le i\le d_F$, if $j\in\mathcal{S}_i$, then by Lemma A.3 there exists a $\theta_A\in\Theta_0(P)$ with $\theta_0^{(j)}\ne\theta_A^{(j)}$. Thus, by convexity of Θ and $P\in\mathbf{P}$ satisfying Assumption 3.6(ii), we obtain that $c\theta_0+(1-c)\theta_A\in\Theta_0(P)$ for all $c\in(0,1)$ and:

$$F^{(i)}(\int m(x,c\theta_0+(1-c)\theta_A)dP(x))<0\ . \tag{A.16}$$

Therefore, (A.15) and (A.16) imply that $\mathcal{S}_i=\emptyset$ for all $i\in\mathcal{A}(\theta_0,P)$, or equivalently that only linear constraints can be active at θ_0. Thus, Theorem 22.2 in Rockafellar (1970) then yields that either (A.17) or (A.18) must hold:

$$F^{(i)}(\int m(x,\theta_L)dP(x))<0 \text{ for all } i\in\mathcal{A}(\theta_0,P) \text{ for some } \theta_L\in\mathbf{R}^{d_\theta} \tag{A.17}$$

$$\sum_{i\in\mathcal{A}(\theta_0,P)}\gamma_i\nabla_\theta\{F^{(i)}(\int m(x,\theta_0)dP(x))\}=0 \text{ for scalars } \{\gamma_i\} \text{ with } \sup_{i\in\mathcal{A}(\theta_0,P)}\gamma_i>0\ . \tag{A.18}$$

However, (A.18) is not possible due to $P\in\mathbf{P}$ satisfying Assumption 3.6(iv) and hence we conclude (A.17) must hold. Finally, since $F^{(i)}(\int m(x,\theta_0)dP(x))<0$ for all $i\in\{1,\ldots,d_F\}\setminus\mathcal{A}(\theta_0,P)$ and $\theta_0\in\Theta^o$ due to $P\in\mathbf{P}$ satisfying Assumption 3.6(i), we obtain that for $c\in(0,1)$ sufficiently close to one $F^{(i)}(\int m(x,c\theta_0+(1-c)\theta_L)dP(x))<0$ for all $1\le i\le d_F$. Hence, (A.15) implies $\mathcal{A}(\theta_0,P)=\emptyset$ as desired, and the claim of the Lemma follows. ■

Lemma A.5. *Let Assumptions 3.2, 3.4(i)-(ii), 3.5, 4.2(i) hold and $P\in\mathbf{P}$. Then, there exists a neighborhood $N(P)\subseteq\mathbf{M}$ such that the mapping $(\theta,Q)\mapsto F(\int m(x,\theta)dQ(x))$ is continuous at all $(\theta,Q)\in\Theta\times N(P)$.*

Proof: Recall that by Lemma A.2 there is $N(P)\subseteq\mathbf{M}$ such that $\int m(x,\theta)dQ(x)\in V_0$ for all $(\theta,Q)\in\Theta\times N(P)$. Next let $\{\theta_\alpha,Q_\alpha\}_{\alpha\in\mathfrak{A}}$ be a net such that $(\theta_\alpha,Q_\alpha)\to(\theta_0,Q_0)\in\Theta\times N(P)$. Since $m:\mathcal{X}\times\Theta\to\mathbf{R}^{d_m}$ is bounded by Assumption 3.4(i), and $\theta\mapsto m(x,\theta)$ is equicontinuous in x by Assumption 3.4(ii) it follows from Lemma A.1 that:

$$\limsup_\alpha\sup_{\theta\in\Theta}\|F(\int m(x,\theta)dQ_\alpha(x))-F(\int m(x,\theta)dQ_0(x))\|=0\ , \tag{A.19}$$

due to F being uniformly continuous on V_0 by Assumption 3.5(ii). Moreover, since $\int m(x,\theta_0)dQ_0(x)\in V_0$ we have

$$F(\int m(x,\theta_\alpha)dQ_0(x))\to F(\int m(x,\theta_0)dQ_0(x)) \tag{A.20}$$

by Assumptions 3.4(i)-(ii) and the dominated convergence theorem. Therefore, results (A.19) and (A.20) imply that

$$F(\int m(x,\theta_\alpha)dQ_\alpha(x))\to F(\int m(x,\theta_0)dQ_0(x))\ , \tag{A.21}$$

which establishes the continuity of $(\theta,Q)\to F(\int m(x,\theta)dQ(x))$ on $\Theta\times N(P)$ as claimed. ■

Corollary A.2. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold and $P\in\mathbf{P}$. Then, there exists a $\theta_0\in\Theta$ and a neighborhood $N(P)\subseteq\mathbf{M}$ such that $F^{(i)}(\int m(x,\theta_0)dQ(x))<0$ for all $1\le i\le d_F$ and $Q\in N(P)$.*

Proof: The claim follows immediately from Lemma A.4 implying there exists $\theta_0 \in \Theta$ such that $F^{(i)}(\int m(x,\theta_0)dP(x)) < 0$ for all $1 \le i \le d_F$, and Lemma A.5 implying $Q \mapsto F(\int m(x,\theta_0)dQ(x))$ is continuous at $Q = P$. ■

Lemma A.6. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold and $P \in \mathbf{P}$. Then there is $N(P) \subseteq \mathbf{M}$ such that $\Theta_0(Q) \neq \emptyset$ is convex for all $Q \in N(P)$, and the correspondence $Q \mapsto \Theta_0(Q)$ is continuous at all $Q \in N(P)$.*

Proof: By Θ being convex, Corollary A.2, Assumption 4.2(i) and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii) there exists a $N(P) \subseteq \mathbf{M}$ and $\theta_0 \in \Theta$ such that for all $Q \in N(P)$ and $1 \le i \le d_F$, the functions $F^{(i)}(\int m(x,\cdot)dQ(x)) : \Theta \to \mathbf{R}$ are convex, and $F^{(i)}(\int m(x,\theta_0)dQ(x)) < 0$. Thus, in what follows we let $\Theta_0(Q)$ be a convex set with nonempty interior. Moreover, by Lemma A.5 $N(P)$ may be chosen so that $(\theta, Q) \mapsto F(\int m(x,\theta)dQ(x))$ is continuous on $\Theta \times N(P)$.

We first establish $Q \mapsto \Theta_0(Q)$ is lower hemicontinuous at any $Q_0 \in N(P)$. By Theorem 17.19 in Aliprantis and Border (2006), it suffices to show that for any $\theta^* \in \Theta_0(Q_0)$ and net $\{Q_\alpha\}_{\alpha \in \mathfrak{A}}$ with $Q_\alpha \to Q_0$, there exists a subnet $\{Q_{\alpha_\beta}\}_{\beta \in \mathfrak{B}}$ and net $\{\theta_\beta\}_{\beta \in \mathfrak{B}}$ such that $\theta_\beta \in \Theta_0(Q_{\alpha_\beta})$ for all $\beta \in \mathfrak{B}$ and $\theta_\beta \to \theta^*$. If $\theta^* \in \Theta_0^o(Q_0)$, then $F^{(i)}(\int m(x,\theta^*)dQ_0(x)) < 0$ for all $1 \le i \le d_F$ and hence by Lemma A.5 and $Q_\alpha \to Q_0$, there exists α_0 such that $\theta^* \in \Theta_0(Q_\alpha)$ for all $\alpha \ge \alpha_0$. Therefore, defining $\mathfrak{B} \equiv \{\alpha \in \mathfrak{A} : \alpha \ge \alpha_0\}$, $Q_{\alpha_\beta} = Q_\beta$ and setting $\theta_\beta = \theta^*$ we obtain $\{Q_{\alpha_\beta}\}_{\beta \in \mathfrak{B}}$ is a subnet with $\theta_\beta \in \Theta_0(Q_{\alpha_\beta})$ and trivially satisfies $\theta_\beta \to \theta^*$. 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 $\tilde{\theta}_k \to \theta^*$ and $\tilde{\theta}_k \in \Theta_0^o(Q_0)$ for all k. By Lemma A.5, there then exits a $\alpha_{0,k}$ such that $\tilde{\theta}_k \in \Theta_0(Q_\alpha)$ for all $\alpha \ge \alpha_{0,k}$. Let $\mathfrak{B} \equiv \mathfrak{A} \times \mathbb{N}$ and for any $\beta = (\alpha, k)$ let $\alpha_\beta = \tilde{\alpha}$ for some $\tilde{\alpha} \in \mathfrak{A}$ with $\tilde{\alpha} \ge \alpha$ and $\tilde{\alpha} \ge \alpha_{0,k}$ and $\theta_\beta = \tilde{\theta}_k$. $\{Q_{\alpha_\beta}\}_{\beta \in \mathfrak{B}}$ is then a subnet of $\{Q_\alpha\}_{\alpha \in \mathfrak{A}}$ with $\theta_\beta \in \Theta_0(Q_{\alpha_\beta})$ and $\theta_\beta \to \theta^*$.

Next, we show that $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous at any $Q_0 \in N(P)$. By Theorem 17.16 in Aliprantis and Border (2006), it suffices to show that any net $\{Q_\alpha, \theta_\alpha\}_{\alpha \in \mathfrak{A}}$ such that $Q_\alpha \to Q_0$ and $\theta_\alpha \in \Theta_0(Q_\alpha)$ for all $\alpha \in \mathfrak{A}$ is such that $\{\theta_\alpha\}_{\alpha \in \mathfrak{A}}$ has a limit point $\theta^* \in \Theta_0(Q_0)$. Compactness of Θ, however, implies there exists a subnet $\{\theta_{\alpha_\beta}\}_{\beta \in \mathfrak{B}}$ such that $\theta_{\alpha_\beta} \to \theta^*$ for some $\theta^* \in \Theta$. Therefore, since $\theta_{\alpha_\beta} \in \Theta_0(Q_{\alpha_\beta})$ for all $\beta \in \mathfrak{B}$, we obtain

$$0 \ge F(\int m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x)) \to F(\int m(x,\theta^*)dQ_0(x)) \tag{A.22}$$

by Lemma A.5. 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 A.3. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold and $P \in \mathbf{P}$. Then, there exists a neighborhood $N(P) \subseteq \mathbf{M}$ such that $\emptyset \neq \Theta_0(Q) \subset \Theta^o$ for all $Q \in N(P)$.*

Proof: Since $\theta \mapsto F(\int m(x,\theta)dP(x))$ is continuous in $\theta \in \Theta$ by Lemma A.5, it follows that $\Theta_0(P)$ is closed. Hence, since $\partial\Theta$ is closed as well and $\Theta_0(P) \cap \partial\Theta = \emptyset$ due to $P \in \mathbf{P}$ satisfying Assumption 3.6(i), we must have that:

$$\inf_{\theta_1 \in \Theta_0(P)} \inf_{\theta_2 \in \partial\Theta} \|\theta_1 - \theta_2\| > 0 \ . \tag{A.23}$$

Therefore, there exists an open set U such that $\Theta_0(P) \subset U \subset \Theta^o$. Since by Lemma A.6 the correspondence $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous at P, there then exists a $N(P) \subseteq \mathbf{M}$ such that for all $Q \in N(P)$ we have $\emptyset \neq \Theta_0(Q) \subset U \subset \Theta^o$; see Definition 17.2 in Aliprantis and Border (2006). ■

Lemma A.7. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold, $P \in \mathbf{P}$ and define the correspondence:*

$$\Xi(p, Q) \equiv \arg\max_{\theta \in \Theta} \{\langle p, \theta \rangle \quad s.t. \ F(\int m(x,\theta)dQ(x)) \le 0\} \ . \tag{A.24}$$

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

Proof: By Lemma A.6, there exists a $N(P) \subseteq \mathbf{M}$ such that $\Theta_0(Q) \neq \emptyset$ and $Q \mapsto \Theta_0(Q)$ is continuous on $N(P)$. Since by Lemma A.5 the set $\Theta_0(Q) \subseteq \Theta$ is closed, Assumption 3.2(i) implies $\Theta_0(Q)$ is compact. Hence, $\Xi(p, Q)$ is well defined as the maximum is indeed attained for all $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P)$. 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 A.8. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold and $P \in \mathbf{P}$. Then, there exists a neighborhood $N(P) \subseteq \mathbf{M}$ 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 $Q \in N(P)$.*

Proof: The proof is by contradiction. Let $\mathfrak{N}_P$ be the neighborhood system of P with direction $V \succeq W$ whenever $V \subseteq W$, which forms a directed set. If the Lemma fails to hold, then for $\mathfrak{A} = \mathfrak{N}_P$ there exists a net $\{Q_\alpha, \theta_\alpha\}_{\alpha\in\mathfrak{A}}$ such that $Q_\alpha \to P$, $\theta_\alpha \in \Theta_0(Q_\alpha)$ 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 $\alpha \in \mathfrak{A}$. Since by Lemma A.6 the correspondence $Q \mapsto \Theta_0(Q)$ is upper hemicontinuous in a neighborhood of P, we may pass to a subnet $\{Q_{\alpha_\beta}, \theta_{\alpha_\beta}\}_{\beta\in\mathfrak{B}}$ such that $(Q_{\alpha_\beta}, \theta_{\alpha_\beta}) \to (P, \theta^*)$ with $\theta^* \in \Theta_0(P)$. Further note that for any index $i \in \mathcal{A}^c(\theta^*, P)$ Lemma A.5 implies that:

$$F^{(i)}(\int m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x)) \to F^{(i)}(\int m(x,\theta^*)dP(x)) < 0 \; . \tag{A.25}$$

Therefore, there is a β_0 such that if $\beta \geq \beta_0$ then the constraints that are inactive under (θ^*, P) are also inactive under $(\theta_{\alpha_\beta}, Q_{\alpha_\beta})$. Equivalently, for $\beta \geq \beta_0$, $\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) \subseteq \mathcal{A}(\theta^*, P)$, and hence in establishing a contradiction it suffices to show $\{\nabla F^{(i)}(\int m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x)) \int \nabla_\theta m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x)\}_{i\in\mathcal{A}(\theta^*,P)}$ are linearly independent for some $\beta \geq \beta_0$.

Towards this end, notice that Assumptions 3.4(ii)-(iii) and Lemma A.1 imply that uniformly in $\theta \in \Theta$:

$$\int \nabla_\theta m(x,\theta)dQ_{\alpha_\beta}(x) \to \int \nabla_\theta m(x,\theta)dP(x) \; . \tag{A.26}$$

Since $\nabla_\theta m$ is uniformly bounded and continuous in θ, the dominated convergence theorem and (A.26) yield:

$$\int \nabla_\theta m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x) \to \int \nabla_\theta m(x,\theta^*)dP(x) \; . \tag{A.27}$$

Similarly, since $v \mapsto \nabla F(v)$ is uniformly continuous on V_0 by Assumption 3.5(ii) and $\int m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x) \in V_0$ for β sufficiently large by Lemma A.2, Lemma A.1 applied to $\theta \mapsto m(x,\theta)$ and result (A.27) yield:

$$\nabla F(\int m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x)) \int \nabla_\theta m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x) \to \nabla F(\int m(x,\theta^*)dP(x)) \int \nabla_\theta m(x,\theta^*)dP(x) \; . \tag{A.28}$$

However, since $P \in \mathbf{P}$ satisfies Assumption 3.6(iv), the vectors $\{\nabla F^{(i)}(\int m(x,\theta^*)dP(x)) \int \nabla_\theta m(x,\theta^*)dP(x)\}_{i\in\mathcal{A}(\theta^*,P)}$ are linearly independent and hence by (A.28), so must $\{\nabla F^{(i)}(\int m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x)) \int \nabla_\theta m(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x)\}_{i\in\mathcal{A}(\theta^*,P)}$ for $\beta \geq \beta_1$ and some $\beta_1 \in \mathfrak{B}$. Thus, the contradiction is established and the claim of the Lemma follows. ■

Lemma A.9. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold and $P \in \mathbf{P}$. Then there is a neighborhood $N(P) \subseteq \mathbf{M}$ such that for all $Q \in N(P)$ and $p \in \mathbb{S}^{d_\theta}$ there is a unique $\lambda(p, Q) \in \mathbf{R}^{d_F}$ satisfying:*

$$\sup_{\theta\in\Theta_0(Q)} \langle p, \theta\rangle = \sup_{\theta\in\Theta} \{\langle p, \theta\rangle + \lambda(p,Q)' F(\int m(x,\theta)dQ(x))\} \; . \tag{A.29}$$

Proof: By Assumption 4.2(i), Corollary A.2 and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii), there is a $N_1(P) \subseteq \mathbf{M}$ such that for all $Q \in N_1(P)$ there is a $\theta_0 \in \Theta$ with $F^{(i)}(\int m(x,\theta_0)dQ(x)) < 0$ for all $1 \leq i \leq d_F$ and $F^{(i)}(\int m(x,\cdot)dQ(x)) : \Theta \to \mathbf{R}$ is convex for all $1 \leq i \leq d_F$. Since Θ is compact and convex by Assumption 3.2(i), the optimization problem:

$$\sup_{\theta\in\Theta} \langle p, \theta\rangle \quad \text{s.t. } F(\int m(x,\theta)dQ(x)) \leq 0 \tag{A.30}$$

satisfies the conditions of Corollary 28.2.1 in Rockafellar (1970) for all $Q \in N_1(P)$ and all $p \in \mathbb{S}^{d_\theta}$. We can therefore conclude that the equality in (A.29) holds for some $\lambda(p, Q) \in \mathbf{R}^{d_F}$.

Next we show there exists a $N(P) \subseteq N_1(P)$ such that $\lambda(p, Q)$ is unique for all $p \in \mathbb{S}^{d_\theta}$ and $Q \in N(P)$. To this end, note that by Lemmas A.7 and Corollary A.3 there exists a $N_2(P) \subseteq N_1(P)$ such that $\Xi(p, Q)$ as defined in

(A.24) satisfies $\emptyset \neq \Xi(p,Q) \subseteq \Theta_0(Q) \subset \Theta^o$ for all $(p,Q) \in \mathbb{S}^{d_\theta} \times N_2(Q)$. Theorem 8.3.1 in Luenberger (1969) then implies that any $\theta^* \in \Xi(p,Q)$ is also a maximizer of the dual problem, and hence for any $\theta^* \in \Xi(p,Q)$:

$$p' + \lambda(p,Q)'\nabla F(\int m(x,\theta^*)dQ(x)) \int \nabla_\theta m(x,\theta^*)dQ(x) = 0 \ , \tag{A.31}$$

by Corollary A.1 for all Q in some neighborhood $N_3(P) \subseteq N_2(P)$. Result (A.31) 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 (A.31) can be reduced to d_θ equations and $\#\mathcal{A}(\theta^*,Q)$ unknowns. Furthermore, by Lemma A.8 there is a neighborhood $N(P) \subseteq N_3(P)$ with $\{\nabla F^{(i)}(\int m(x,\theta^*)dQ(x)) \int \nabla_\theta m(x,\theta^*)dQ(x)\}_{i\in\mathcal{A}(\theta^*,Q)}$ linearly independent for all $Q \in N(P)$ and any $\theta^* \in \Theta_0(Q)$. Hence, we conclude that for any $Q \in N(P)$ the solution to equation (A.31) in $\lambda(p,Q) \in \mathbf{R}^{d_F}$ satisfying (A.30) is unique and the claim of the Lemma follows. ■

Lemma A.10. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold, $P \in \mathbf{P}$, and $\Xi(p,Q)$ be as in* (A.24)*. Then, there exists a neighborhood $N(P) \subseteq \mathbf{M}$ such for each $(Q,p) \in N(P) \times \mathbb{S}^{d_\theta}$ and all $1 \le i \le d_F$ one of the following must hold: (i) $\lambda^{(i)}(p,Q) = 0$, or (ii) $\theta_1^{(j)} = \theta_2^{(j)}$ for all $j \in \mathcal{S}_i$ and all $\theta_1,\theta_2 \in \Xi(p,Q)$.*

Proof: Recall we refer to the *arguments* of $F_S^{(i)}(\int m_S(x,\cdot)dQ(x))$ as the coordinates of θ corresponding to indices in $\mathcal{S}_i$ (as in (4)). By $P \in \mathbf{P}$ satisfying Assumption 3.6(ii) and Lemma A.7, there is a $N(P) \subseteq \mathbf{M}$ such that for all $Q \in N(P)$ and $1 \le i \le d_F$, the functions $F_S^{(i)}(\int m_S(x,\cdot)dQ(x))$ are strictly convex in their arguments, and $\Xi(p,Q) \neq \emptyset$ for all $p \in \mathbb{S}^{d_\theta}$. To establish the Lemma, we aim to show that condition (i) must hold whenever (ii) fails. To this end, suppose there exist a $1 \le i \le d_F$ such that $\theta_1^{(j)} \neq \theta_2^{(j)}$ for some $j \in \mathcal{S}_i$ and $\theta_1,\theta_2 \in \Xi(p,Q)$. Next define $\theta_L = c\theta_1 + (1-c)\theta_2$ with $c \in (0,1)$ and note $\theta_1^{(j)} \neq \theta_2^{(j)}$ and $j \in \mathcal{S}_i$, and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii) imply

$$F^{(i)}(\int m(x,\theta_L)dQ(x)) < cF^{(i)}(\int m(x,\theta_1)dQ(x)) + (1-c)F^{(i)}(\int m(x,\theta_2)dQ(x)) \le 0 \tag{A.32}$$

where the second inequality follows from $\theta_1,\theta_2 \in \Theta_0(Q)$. However, since Θ is convex by Assumption 3.2(i), $\Theta_0(Q)$ is convex as well and hence $\theta_L \in \Theta_0(Q)$. Since $\langle p,\theta_L\rangle = c\langle p,\theta_1\rangle + (1-c)\langle p,\theta_2\rangle$, we must have $\theta_L \in \Xi(p,Q)$, and therefore (A.32) and the complementary slackness condition imply $\lambda^{(i)}(p,Q) = 0$, establishing the Lemma. ■

Lemma A.11. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold, $P \in \mathbf{P}$, and $\lambda(p,Q)$ be as in* (A.29)*. Then, there exists a $N(P) \subseteq \mathbf{M}$ such that $\|\lambda(p,Q)\|$ is uniformly bounded in $(p,Q) \in \mathbb{S}^{d_\theta} \times N(P)$.*

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

$$\|\lambda(p_\alpha,Q_\alpha)\| > k \qquad Q_\alpha \in V \qquad p_\alpha \in \mathbb{S}^{d_\theta} \qquad \theta_\alpha \in \Xi(p_\alpha,Q_\alpha) \ , \tag{A.33}$$

where $\Xi(p,Q)$ is as in (A.24). However, by: (i) $(p,Q) \mapsto \Xi(p,Q)$ being upper hemicontinuous and compact valued in a neighborhood of P, and (ii) $\mathbb{S}^{d_\theta}$ being compact, we may pass to a subnet $\{Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}\}_{\beta\in\mathfrak{B}}$ such that:

$$(Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}, \|\lambda(p_{\alpha_\beta},Q_{\alpha_\beta})\|) \to (P, p^*, \theta^*, +\infty) \ \text{ for some } \ (p^*,\theta^*) \in \mathbb{S}^{d_\theta} \times \Xi(p^*,P) \ . \tag{A.34}$$

Since the number of constraints is finite, there is a set of indices $\mathcal{C} \subseteq \{1,\ldots,d_F\}$ such that for every $\beta_0 \in \mathfrak{B}$ there exists a $\beta \ge \beta_0$ with $\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) = \mathcal{C}$. Letting $\mathfrak{G} \equiv \mathfrak{B}$ we may then set $\alpha_{\beta_\gamma} = \alpha_{\tilde\beta}$ for some $\tilde\beta \ge \beta$ 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 $\{Q_{\alpha_\gamma}, p_{\alpha_\gamma}, \theta_{\alpha_\gamma}\}_{\gamma\in\mathfrak{G}}$ with:

$$(Q_{\alpha_\gamma}, p_{\alpha_\gamma}, \theta_{\alpha_\gamma}, \|\lambda(p_{\alpha_\gamma},Q_{\alpha_\gamma})\|) \to (P, p^*, \theta^*, +\infty) \qquad \mathcal{A}(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) = \mathcal{C} \ \ \forall \gamma \in \mathfrak{G} \ . \tag{A.35}$$

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 $\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})$ and $\nabla F(\int m(x,\theta_{\alpha_\gamma})dQ_{\alpha_\gamma}(x))$ whose indexes belong to $\mathcal{C}$. Similarly, define:

$$M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) \equiv \nabla^{\mathcal{C}} F(\int m(x,\theta_{\alpha_\gamma})dQ_{\alpha_\gamma}(x)) \int \nabla_\theta m(x,\theta_{\alpha_\gamma})dQ_{\alpha_\gamma}(x) \ . \tag{A.36}$$

By Lemma A.8 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 (A.31) that:

$$\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma}) = -(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})')^{-1}M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})p_{\alpha_\gamma} \ . \tag{A.37}$$

Additionally, since $(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma}) \to (\theta^*, P)$ as in (A.34), we obtain from result (A.28) and definition (A.36) that:

$$M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})' \to M(\theta^*, P)M(\theta^*, P)' \ . \tag{A.38}$$

For a symmetric matrix Σ, let $\xi(\Sigma)$ denote its smallest eigenvalue and note $\xi(M(\theta^*, P)M(\theta^*, P)') > 2\epsilon$ for some $\epsilon > 0$ by $P \in \mathbf{P}$ satisfying Assumption 3.6(iv). Since eigenvalues are continuous under $\|\cdot\|_F$ by Corollary III.2.6 in Bhatia (1997), we obtain from (A.38) that there is a $\gamma_1 \geq \gamma_0 \in \mathfrak{G}$ such that for all $\gamma \geq \gamma_1$ we have

$$\xi(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})') > \epsilon \ . \tag{A.39}$$

Furthermore, since $\lambda^{(i)}(p_{\alpha_\gamma}, Q_{\alpha_\gamma}) = 0$ for all $i \notin \mathcal{C}$, it follows that $\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\| = \|\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\|$ and hence:

$$\begin{aligned}\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\| &= \|\lambda^{\mathcal{C}}(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\| \\ &\leq \|(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})')^{-1}\|_o \times \|M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})\|_F \times \|p\| \\ &\leq \xi^{-1}(M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})M(\theta_{\alpha_\gamma}, Q_{\alpha_\gamma})') \times \sup_{v \in V_0} \|\nabla F(v)\|_F \times \sup_{(x,\theta) \in \mathcal{X} \times \Theta} \|\nabla_\theta m(x, \theta)\|_F \ ,\end{aligned} \tag{A.40}$$

where the final inequality holds for all $\gamma \geq \gamma_2$ for some $\gamma_2 \in \mathfrak{G}$ with $\gamma_2 \geq \gamma_1$ by Lemma A.2. However, (A.39), (A.40) and Assumptions 3.4(ii), 3.5(ii) imply $\|\lambda(p_{\alpha_\gamma}, Q_{\alpha_\gamma})\|$ is uniformly bounded for all $\gamma \geq \gamma_2$, contradicting (A.35). ■

Lemma A.12. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold, $P \in \mathbf{P}$ and $\lambda(p, Q)$ be as in* (A.29)*. Then, there exists a $N(P) \subseteq \mathbf{M}$ such that the function $(p, Q) \mapsto \lambda(p, Q)$ is continuous on $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P)$.*

Proof: By Lemmas A.9 and A.11 there exists a $N_1(P) \subseteq \mathbf{M}$ such that $\lambda(p, Q)$ is well defined, unique and uniformly bounded for all $(p, Q) \in \mathbb{S}^{d_\theta} \times N_1(P)$. Therefore, letting $\Lambda \equiv \text{cl}\{\lambda(p, Q) : (p, Q) \in \mathbb{S}^{d_\theta} \times N_1(P)\}$ it follows that Λ is compact in $\mathbf{R}^{d_F}$. By Lemma A.9 and Theorem 8.6.1 in Luenberger (1969) we then have:

$$\lambda(p, Q) = \arg\min_{\lambda \geq 0} V(\lambda, p, Q) = \arg\min_{\lambda \in \Lambda} V(\lambda, p, Q) \qquad V(\lambda, p, Q) \equiv \max_{\theta \in \Theta}\{\langle p, \theta\rangle + \lambda' F(\int m(x, \theta) dQ(x))\} \ . \tag{A.41}$$

Since $(\theta, Q) \mapsto F(\int m(x, \theta) dQ(x))$ is continuous on a neighborhood $N(P) \subseteq N_1(P)$ by Lemma A.5, compactness of Θ and Theorem 17.31 in Aliprantis and Border (2006) imply $(\lambda, p, Q) \mapsto V(\lambda, p, Q)$ is continuous on $\Lambda \times \mathbb{S}^{d_\theta} \times N(P)$. Therefore, by (A.41), 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 on $\mathbb{S}^{d_\theta} \times N(P)$. However, since $(p, Q) \mapsto \lambda(p, Q)$ is a singleton valued correspondence on $\mathbb{S}^{d_\theta} \times N(P)$ by Lemma A.9, we conclude that it is in fact a continuous function. ■

Lemma A.13. *Let Assumptions 3.2, 3.4, 3.5, 4.2(i) hold, $P \in \mathbf{P}$ and $\Xi(p, P)$ be as in* (A.24)*. Then, there exists a Borel measurable selector $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ with $\theta^*(p) \in \Xi(p, P)$ for all $p \in \mathbb{S}^{d_\theta}$.*

Proof: By Lemma A.7, $p \mapsto \Xi(p, P)$ is upper hemicontinuous in $p \in \mathbb{S}^{d_\theta}$ and hence weakly measurable; see Definition 18.1 in Aliprantis and Border (2006). Since $p \mapsto \Xi(p, P)$ is nonempty and compact valued by Lemma A.7, Theorem 18.13 in Aliprantis and Border (2006), implies there is a measurable selector $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ and the Lemma follows. ■

Lemma A.14. *Let Assumptions 3.2, 3.3, 3.4, 3.5 hold, and $\eta \mapsto h_\eta$ be a curve in $\mathbf{S}$. Then, there is a neighborhood $N \subseteq \mathbf{R}$ of 0 such that for all $\eta_0 \in N$, $p \in \mathbb{S}^{d_\theta}$, $\Xi(p, P_\eta)$ as in* (A.24) *and $\lambda(p, P_\eta) \in \mathbf{R}^{d_F}$ as in* (A.29),

$$\begin{aligned}&\frac{\partial}{\partial \eta}\nu(p, \Theta_0(P_\eta))\Big|_{\eta = \eta_0} \\ &= 2\lambda(p, P_{\eta_0})' \nabla F(\int m(x, \theta^*) h_{\eta_0}^2(x) d\mu(x)) \int m(x, \theta^*) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \quad \textit{for any } \theta^* \in \Xi(p, P_{\eta_0}) \ .\end{aligned} \tag{A.42}$$

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:

$$|\int m^{(i)}(x,\theta)\{h_{\eta_0}^2(x) - h_\eta^2(x) - 2(\eta_0 - \eta)h_{\eta_0}(x)\dot{h}_{\eta_0}(x)\}d\mu(x)|$$
$$= |\int m^{(i)}(x,\theta)\{(h_\eta(x) - h_{\eta_0}(x))^2 + 2h_{\eta_0}(x)(h_\eta(x) - h_{\eta_0}(x) + (\eta_0 - \eta)\dot{h}_{\eta_0}(x))\}d\mu(x)| = o(|\eta - \eta_0|) \quad \text{(A.43)}$$

where the final result holds by m being bounded by Assumption 3.4(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. Moreover, $\|h_\eta - h_{\eta_0}\|_{L_\mu^2} = o(1)$ implies $P_\eta \to P_{\eta_0}$ 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 (A.43), Lemma A.2 and Assumptions 3.5(i)-(ii) yield:

$$\frac{\partial}{\partial \eta} F(\int m(x,\theta)h_\eta^2(x)d\mu(x))\Big|_{\eta=\eta_0} = 2\nabla F(\int m(x,\theta)h_{\eta_0}^2(x)d\mu(x)) \int m(x,\theta)\dot{h}_{\eta_0}(x)h_{\eta_0}(x)d\mu(x) \ . \quad \text{(A.44)}$$

Since $\eta \mapsto h_\eta$ is continuously Fréchet differentiable, result (A.28) implies the derivative in (A.44) is continuous in η_0 in a neighborhood of zero. Therefore, Assumption 3.3(i) implying Assumption 4.2(i), Lemma A.9 and Corollary 5 in Milgrom and Segal (2002) imply $\eta \mapsto \nu(p, \Theta_0(P_\eta))$ is directionally differentiable in a neighborhood of zero with:

$$\frac{\partial}{\partial \eta_+}\nu(p,\Theta_0(P_\eta))\Big|_{\eta=\eta_0} = \max_{\theta^* \in \Xi(p,P_{\eta_0})} 2\lambda(p,P_{\eta_0})'\nabla F(\int m(x,\theta^*)h_{\eta_0}^2(x)d\mu(x)) \int m(x,\theta^*)\dot{h}_{\eta_0}(x)h_{\eta_0}(x)d\mu(x) \quad \text{(A.45)}$$
$$\frac{\partial}{\partial \eta_-}\nu(p,\Theta_0(P_\eta))\Big|_{\eta=\eta_0} = \min_{\theta^* \in \Xi(p,P_{\eta_0})} 2\lambda(p,P_{\eta_0})'\nabla F(\int m(x,\theta^*)h_{\eta_0}^2(x)d\mu(x)) \int m(x,\theta^*)\dot{h}_{\eta_0}(x)h_{\eta_0}(x)d\mu(x) \quad \text{(A.46)}$$

where $\frac{\partial}{\partial \eta_+}$ and $\frac{\partial}{\partial \eta_-}$ denote right and left derivatives respectively. Note, however, that by Lemma A.10, for all $1 \leq i \leq d_F$ such that $\lambda^{(i)}(p, P_{\eta_0}) \neq 0$ we must have $\theta_1^{(j)} = \theta_2^{(j)}$ for all $j \in \mathcal{S}_i$ and all $\theta_1, \theta_2 \in \Xi(p, P_{\eta_0})$. Therefore, since $A\theta$ trivially does not depend on η, it follows from (3), (4), and results (A.44), (A.45) and (A.46) that:

$$\frac{\partial}{\partial \eta_+}\nu(p,\Theta_0(P))\Big|_{\eta=\eta_0} = \max_{\theta^* \in \Xi(p,P_{\eta_0})} \sum_{i:\lambda^{(i)}(p,P_{\eta_0})\neq 0} \lambda^{(i)}(p,P_{\eta_0})\frac{\partial}{\partial \eta}F_S^{(i)}(\int m(x,\theta^*)h_\eta^2(x)d\mu(x))\Big|_{\eta=\eta_0}$$
$$= \min_{\theta^* \in \Xi(p,P_{\eta_0})} \sum_{i:\lambda^{(i)}(p,P_{\eta_0})\neq 0} \lambda^{(i)}(p,P_{\eta_0})\frac{\partial}{\partial \eta}F_S^{(i)}(\int m(x,\theta^*)h_\eta^2(x)d\mu(x))\Big|_{\eta=\eta_0} = \frac{\partial}{\partial \eta_-}\nu(p,\Theta_0(P))\Big|_{\eta=\eta_0} \ . \quad \text{(A.47)}$$

Thus, the claim of the Lemma follows from (A.45), (A.46) and (A.47). ■

Lemma A.15. *Let Assumptions 3.2, 3.3, 3.4, 3.5 hold, and $\eta \mapsto h_\eta$ be a curve in $\mathbf{S}$. Then: (i) There is a neighborhood $N \subseteq \mathbf{R}$ of 0 such that $\frac{\partial}{\partial \eta}\nu(p,\Theta_0(P_\eta))|_{\eta=\eta_0}$ is bounded in $(p,\eta_0) \in \mathbb{S}^{d_\theta} \times N$, and (ii) The function $(p,\eta_0) \mapsto \frac{\partial}{\partial \eta}\nu(p,\Theta_0(P_\eta))|_{\eta=\eta_0}$ is continuous at all $(p,\eta_0) \in \mathbb{S}^{d_\theta} \times N$.*

Proof: To establish the first claim, notice that by Lemmas A.2, A.14 and the Cauchy-Schwarz inequality:

$$|\frac{\partial}{\partial \eta}\nu(p,\Theta_0(P_\eta))\Big|_{\eta=\eta_0}| \leq 2\|\lambda(p,P_{\eta_0})\| \times \sup_{v \in V_0}\|\nabla F(v)\|_F \times \sqrt{d_m} \sup_{(x,\theta)\in\mathcal{X}\times\Theta}\|m(x,\theta)\| \times \|\dot{h}_{\eta_0}\|_{L_\mu^2} \times \|h_{\eta_0}\|_{L_\mu^2} \ , \quad \text{(A.48)}$$

for η_0 in a neighborhood of zero. Since $\|\dot{h}_{\eta_0}\|_{L_\mu^2}$ is continuous in η_0 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 $\|h_{\eta_0}\|_{L_\mu^2} = 1$ for all η_0, Lemma A.11, Assumptions 3.4(i), 3.5(ii) and (A.48) establish the first claim of the Lemma.

To establish the second claim, let $(p_n, \eta_n) \to (p_0, \eta_0)$ and select $\theta_n^* \in \Xi(p_n, P_{\eta_n})$ for $\Xi(p, P)$ as in (A.24). Since $\|m(x,\theta)\|$ is uniformly bounded by Assumption 3.4(i), we obtain for any $1 \leq i \leq d_m$ that:

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}|\int m^{(i)}(x,\theta)\{\dot{h}_{\eta_n}(x)h_{\eta_n}(x) - \dot{h}_{\eta_0}(x)h_{\eta_0}(x)\}d\mu(x)|$$
$$\leq \sup_{(x,\theta)\in\mathcal{X}\times\Theta}\|m(x,\theta)\| \times \lim_{n\to\infty}\{\|\dot{h}_{\eta_n} - \dot{h}_{\eta_0}\|_{L_\mu^2}\|h_{\eta_n}\|_{L_\mu^2} + \|h_{\eta_n} - h_{\eta_0}\|_{L_\mu^2}\|\dot{h}_{\eta_0}\|_{L_\mu^2}\} = 0 \ , \quad \text{(A.49)}$$

due to the Cauchy-Schwarz inequality, $\eta \mapsto h_\eta$ being continuously Fréchet differentiable and $\|h_\eta\|_{L_\mu^2} = 1$. Next, let $\{n_k\}$ be an arbitrary subsequence, and note that since Lemma A.7 implies $(p,\eta) \mapsto \Xi(p, P_\eta)$ is upper hemicontinuous provided η is in a neighborhood of zero, there is a further subsequence $\{\theta^*_{n_{k_j}}\}$ such that $\theta^*_{n_{k_j}} \to \theta^*$ for some $\theta^* \in \Xi(p_0, P_{\eta_0})$. Along such a subsequence, we obtain from (A.28), (A.49) and the dominated convergence theorem:

$$\lim_{j\to\infty} \nabla F(\int m(x, \theta^*_{n_{k_j}}) h^2_{\eta_{n_{k_j}}}(x) d\mu(x)) \int m(x, \theta^*_{n_{k_j}}) \dot{h}_{\eta_n}(x) h_{\eta_{n_{k_j}}}(x) d\mu(x)$$
$$= \nabla F(\int m(x, \theta^*) h^2_{\eta_0}(x) d\mu(x)) \int m(x, \theta^*) \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \ . \quad \text{(A.50)}$$

Hence, by Lemmas A.12 and A.14 and result (A.50), the subsequence $\{n_k\}$ has a further subsequence $\{n_{k_j}\}$ with:

$$\lim_{j\to\infty} \frac{\partial}{\partial\eta} \nu(p_{n_{k_j}}, \Theta_0(P_\eta))\Big|_{\eta=\eta_{n_{k_j}}} = \frac{\partial}{\partial\eta} \nu(p_0, \Theta_0(P_\eta))\Big|_{\eta=\eta_0} \ . \quad \text{(A.51)}$$

Therefore, since the subsequence $\{n_k\}$ was arbitrary, result (A.51) must also hold with $\{n\}$ in place of $\{n_{k_j}\}$. We conclude that $(p, \eta_0) \mapsto \frac{\partial}{\partial\eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_0}$ is continuous, and the second claim of the Lemma then follows. ■

Lemma A.16. *Let* $\mathbf{M}_\mu \equiv \{Q \in \mathbf{M} : Q \ll \mu\}$, $\mathbf{Q} \subseteq \mathbf{M}_\mu$ *and* $\mathbf{D} \equiv \{s \in L^2_\mu : s = \sqrt{dQ/d\mu}$ *for some* $Q \in \mathbf{Q}\}$. *If* $\mathbf{Q}$ *is open relative to* $\mathbf{M}_\mu$ *with respect to the* τ*-topology, then for every* $Q \in \mathbf{Q}$ *the tangent space of* $\mathbf{D}$ *at* $s = \sqrt{dQ/d\mu}$ *is given by* $\dot{\mathbf{D}} = \{h \in L^2_\mu : \int h(x)s(x)d\mu(x) = 0\}$.

Proof: The proof exploits a construction in Example 3.2.1 of Bickel et al. (1993). Define:

$$\mathbf{T} \equiv \{h \in L^2_\mu : \int h(x)s(x)d\mu(x) = 0\} \ , \quad \text{(A.52)}$$

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

$$h^2_\eta(x) \equiv b(\eta)s^2(x)\Psi\Big(\frac{2\eta h(x)}{s(x)}\Big) \qquad b(\eta) \equiv \Big[\int \Psi\Big(\frac{2\eta h(x)}{s(x)}\Big)dQ(x)\Big]^{-1} \ . \quad \text{(A.53)}$$

Employing Proposition 2.1.1 in Bickel et al. (1993) it is straightforward to verify $\eta \mapsto h_\eta$ is a curve in L^2_μ such that $h_0 = s$. Further note that since $\mathbf{Q}$ is open relative to $\mathbf{M}_\mu$ there exists a neighborhood $N(Q) \subseteq \mathbf{M}$ in the τ-topology such that $N(Q) \cap \mathbf{M}_\mu \subseteq \mathbf{Q}$. Let Q_η satisfy $h_\eta = \sqrt{dQ_\eta/d\mu}$ and notice $2^{-\frac{1}{2}}\|h_\eta - s\|_{L^2_\mu}$ equals the Hellinger distance between Q_η and Q. Since convergence with respect to the Hellinger distance implies convergence with respect to the τ-topology, it follows that there is a neighborhood $N \subseteq \mathbf{R}$ of 0 such that $Q_\eta \in N(Q) \cap \mathbf{M}_\mu \subseteq \mathbf{Q}$ for all $\eta \in N$. We conclude $\eta \mapsto h_\eta$ is a regular parametric submodel. Moreover, by direct calculation we also have:

$$\dot{h}_0(x) = \frac{1}{2}\frac{b(0)s^2(x)\Psi'(0)2h(x)}{s(x)s(x)} + \frac{1}{2}\frac{b'(0)s^2(x)\Psi(0)}{s(x)} = h(x) \ , \quad \text{(A.54)}$$

where we have exploited that by the dominated convergence theorem $b'(0) = 2\int \Psi'(0)h(x)s(x)d\mu(x) = 0$ due to $h \in \mathbf{T}$. Hence, from (A.54) we conclude that $h \in \dot{\mathbf{D}}$ and therefore that $\mathbf{T} = \dot{\mathbf{D}}$, which establishes the Lemma. ■

Theorem A.1. *Let Assumptions 3.2, 3.3, 3.4, 3.5, hold and* $P \in \mathbf{P}$. *Then, the tangent space of* $\mathbf{S}$ *at* $s \equiv \sqrt{dP/d\mu}$ *is given by* $\dot{\mathbf{S}} = \{h \in L^2_\mu : \int h(x)s(x)d\mu(x) = 0\}$.

Proof: The claim follows from Assumption 3.3(i) implying 4.2(i), Lemma A.16 and Lemmas A.2, A.8, Corollary A.3, and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii) implying that $\mathbf{P}$ is open in $\mathbf{M}_\mu \equiv \{Q \in \mathbf{M} : Q \ll \mu\}$. ■

Theorem A.2. *If Assumptions 3.2, 3.3, 3.4 and 3.5 hold, then the mapping* $\rho : \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ *pointwise defined by* $\rho(P) = \nu(\cdot, \Theta_0(P))$ *is weak-pathwise differentiable at any* $P \in \mathbf{P}$. *Moreover, for* $s \equiv \sqrt{dP/d\mu}$ *and* $\lambda(p, Q)$ *as defined in* (A.29), *the derivative* $\dot{\rho} : \dot{\mathbf{S}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ *satisfies:*

$$\dot{\rho}(\dot{h}_0)(p) = 2\lambda(p, P)'\nabla F(\int m(x, \theta^*(p))dP(x)) \int m(x, \theta^*(p))\dot{h}_0(x)s(x)d\mu(x) \ ,$$

where $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ *is Borel measurable and satisfies* $\theta^*(p) \in \Xi(p, P)$ *(as in* (A.24)*) for all* $p \in \mathbb{S}^{d_\theta}$.

Proof: The existence of a Borel measurable $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ satisfying $\theta^*(p) \in \Xi(p, P)$ for all $p \in \mathbb{S}^{d_\theta}$ follows from Lemma A.13. Moreover, notice that indeed $\dot{\rho}(\dot{h}_0) \in \mathcal{C}(\mathbb{S}^{d_\theta})$ for all $\dot{h}_0 \in \dot{\mathbf{S}}$ as implied by Lemmas A.14 and A.15. We next establish that $\dot{\rho} : \dot{\mathbf{S}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ is a continuous linear operator and then verify it is indeed the derivative of $\rho : \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$. Linearity is immediate, while continuity follows by noting that by the Cauchy-Schwarz inequality:

$$\sup_{\|\dot{h}_0\|_{L^2_\mu}=1} \|\dot{\rho}(\dot{h}_0)\|_\infty$$
$$\leq \sup_{\|\dot{h}_0\|_{L^2_\mu}=1} \sup_{p \in \mathbb{S}^{d_\theta}} \{2\|\lambda(p,P)\| \times \sup_{v \in V_0} \|\nabla F(v)\|_F \times \sqrt{d_m} \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \|m(x,\theta)\| \times \|\dot{h}_0\|_{L^2_\mu} \times \|s\|_{L^2_\mu}\} < \infty \ , \quad \text{(A.55)}$$

where we exploited $P \in \mathbf{P}$ satisfies Assumption 3.6(iii), Lemma A.11, Assumptions 3.4(i), 3.5(ii) and $\|s\|_{L^2_\mu} = 1$.

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

$$\lim_{\eta_0 \to 0} \int_{\mathbb{S}^{d_\theta}} \{\frac{\nu(p, \Theta_0(P_{\eta_0})) - \nu(p, \Theta_0(P))}{\eta_0} - \dot{\rho}(\dot{h}_0)(p)\} dB(p) = 0 \quad \text{(A.56)}$$

for all curves $\eta \mapsto P_\eta$ in $\mathbf{P}$ with $h_0 = s$ and all finite Borel measures B on $\mathbb{S}^{d_\theta}$. However, by the mean value theorem:

$$\lim_{\eta_0 \to 0} \int_{\mathbb{S}^{d_\theta}} \frac{\nu(p, \Theta_0(P_{\eta_0})) - \nu(p, \Theta_0(P))}{\eta_0} dB(p) = \lim_{\eta_0 \to 0} \int_{\mathbb{S}^{d_\theta}} \frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta))\Big|_{\eta = \bar{\eta}(p,\eta_0)} dB(p)$$
$$= \int_{\mathbb{S}^{d_\theta}} \frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta))\Big|_{\eta=0} dB(p) = \int_{\mathbb{S}^{d_\theta}} \dot{\rho}(\dot{h}_0)(p) dB(p) \ , \quad \text{(A.57)}$$

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 A.15 justifying the use of the dominated convergence theorem, while the final equality follows by Lemma A.14 and the definition of $\dot{\rho} : \dot{\mathbf{S}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$. Therefore, from (A.57), (A.56) is established. ■

Proof of Theorem 3.2: We employ the framework in Chapter 5.2 in Bickel et al. (1993). Let $\mathbf{B} \equiv \mathcal{C}(\mathbb{S}^{d_\theta})$ and $\mathbf{B}^*$ denote the set of finite Borel measures on $\mathbb{S}^{d_\theta}$, which by Corollary 14.15 in Aliprantis and Border (2006) is the dual space of $\mathbf{B}$. Let $s \equiv \sqrt{dP/d\mu}$ and $\rho : \mathbf{P} \to \mathbf{B}$ be pointwise given by $\rho(P) \equiv \nu(\cdot, \Theta_0(P))$, which has pathwise weak derivative $\dot{\rho}$ at P by Theorem A.2. For $p \mapsto \theta^*(p)$ as in Lemma A.13 and any $B \in \mathbf{B}^*$ then let:

$$\dot{\rho}^T(B)(x) \equiv \int_{\mathbb{S}^{d_\theta}} 2\lambda(p,P)' H(\theta^*(p))\{m(x, \theta^*(p)) - E[m(X_i, \theta^*(p))]\} s(x) dB(p) \ . \quad \text{(A.58)}$$

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

By Assumption 3.4(ii), Lemma A.13 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 $p \mapsto E[m(X_i, \theta^*(p))]$. Similarly, $p \mapsto H(\theta^*(p))$ is measurable by continuity of $\theta \mapsto H(\theta)$ (see (A.28)) and Lemma A.13, while $p \mapsto \lambda(p, P)$ and $x \mapsto s(x)$ are trivially measurable by Lemma A.12 and $s \in L^2_\mu$. The joint measurability of $(p, x) \mapsto (\lambda(p, P), H(\theta^*(p)), m(x, \theta^*(p)), E[m(X_i, \theta^*(p))], s(x))$ 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:

$$(p, x) \mapsto 2\lambda(p,P)' H(\theta^*(p))\{m(x, \theta^*(p)) - E[m(X_i, \theta^*(p))]\} s(x) \quad \text{(A.59)}$$

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 B, $P \in \mathbf{P}$, Lemma A.11 and $\int s^2(x) d\mu(x) = 1$ imply:

$$\int_{\mathcal{X}} (\dot{\rho}^T(B)(x))^2 d\mu(x) \leq \sup_{p \in \mathbb{S}^{d_\theta}} 16\|\lambda(p,P)\|^2 \times \sup_{v \in V_0} \|\nabla F(v)\|_F^2 \times \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \|m(x,\theta)\|^2 \times |B|(\mathbb{S}^{d_\theta}) < \infty \ , \quad \text{(A.60)}$$

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

$$\int_{\mathcal{X}} \dot{\rho}^T(B)(x) s(x) d\mu(x) = 2 \int_{\mathcal{X}} \int_{\mathbb{S}^{d_\theta}} \lambda(p,P)' H(\theta^*(p))\{m(x, \theta^*(p)) - E[m(X_i, \theta^*(p))]\} dB(p) dP(x) = 0 \ . \quad \text{(A.61)}$$

Therefore, by Theorem A.1 and (A.61) we conclude $\dot{\rho}^T(B) \in \dot{\mathbf{S}}$ for all $B \in \mathbf{B}^*$. In addition, we note that since

$$\int_{\mathbb{S}^{d_\theta}} \dot{\rho}(h)(p)dB(p) = \int_{\mathcal{X}} h(x)\dot{\rho}^T(B)(x)d\mu(x) \tag{A.62}$$

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

Finally note Theorem A.1, Theorem A.2 and Theorem 5.2.1 in Bickel et al. (1993) yield:

$$\begin{aligned}\text{Cov}(\int_{\mathbb{S}^{d_\theta}} \mathbb{G}(p)dB_1(p), \int_{\mathbb{S}^{d_\theta}} \mathbb{G}(q)dB_2(q)) &= \frac{1}{4}\int_{\mathcal{X}} \dot{\rho}^T(B_1)(x)\dot{\rho}^T(B_2)(x)d\mu(x) \\ &= \int_{\mathbb{S}^{d_\theta}}\int_{\mathbb{S}^{d_\theta}} \lambda(p,P)'H(\theta^*(p))\Omega(\theta^*(p),\theta^*(q))H(\theta^*(q))'\lambda(q,P)dB_1(p)dB_2(q) \end{aligned} \tag{A.63}$$

for any $B_1, B_2 \in \mathbf{B}^*$, with the second equality following from $s^2 = dP/d\mu$ and reversing the order of integration. Letting B_1 and B_2 equal the degenerate probability measures at p_1 and p_2 in (A.63) then concludes the proof. ■

Appendix B - Proof of Theorems 4.1, 4.2 and Corollary 4.1

In this Appendix we establish Theorems 4.1 and 4.2. The proofs of Theorem 4.2 and Corollary 4.1 are self contained. The proof of Theorem 4.1, however, requires multiple steps, which we outline below.

Step 1: We first establish $\hat{P}_n$ is consistent for P under the τ-topology (Lemma B.5), and that each neighborhood in the τ-topology contains a convex open set (Lemma B.2), which will enable us to employ the mean value theorem. ■

Step 2: Lemma B.3 shows the support function is appropriately differentiable at P, which will enable us to establish:

$$\sqrt{n}\{\nu(p,\Theta_0(\hat{P}_n)) - \nu(p,\Theta_0(P))\} = \sqrt{n}\lambda(p,\hat{P}_{n,\tau_0(p)})'\nabla F(\int m(x,\tilde{\theta}(p))d\hat{P}_{n,\tau_0(p)}(x))\int m(x,\tilde{\theta}(p))(d\hat{P}_n(x) - dP(x))$$

by the mean value theorem, where $\hat{P}_{n,\tau} = \tau\hat{P}_n + (1-\tau)P$, $\tau_0 : \mathbb{S}^{d_\theta} \to [0,1]$ and $\tilde{\theta}(p) \in \Xi(p, \hat{P}_{n,\tau_0(p)})$ for all $p \in \mathbb{S}^{d_\theta}$. ■

Step 3: In Lemma B.8 we exploit equicontinuity (Lemma B.1) to further show that uniformly in $p \in \mathbb{S}^{d_\theta}$

$$\begin{aligned}&\sqrt{n}\lambda(p,P)'\nabla F(\int m(x,\tilde{\theta}(p))d\hat{P}_{n,\tau_0(p)}(x))\int m(x,\tilde{\theta}(p))(d\hat{P}_n(x) - dP(x)) \\ &\quad = \sqrt{n}\lambda(p,P)'\nabla F(\int m(x,\theta^*(p))dP(x))\int m(x,\theta^*(p))(d\hat{P}_n(x) - dP(x)) + o_p(1) ,\end{aligned}$$

where $\theta^*(p) \in \Xi(p,P)$. A key complication is that $\Xi(p,P)$ and $\Xi(p,\hat{P}_{n,\tau_0(p)})$ may not be singleton valued. This problem is addressed employing Lemmas B.4 and B.7. ■

Step 4: Lemma B.9 then verifies Theorem 4.1(ii) using Steps 1, 2 and 3, and continuity of $Q \mapsto \lambda(p,Q)$. Theorem 4.1(iii) is immediate from Lemma B.9 and Lemma B.10, which shows stochastic equicontinuity. ■

Lemma B.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 $\mathcal{F} \equiv \{f : \mathcal{X} \times \mathbf{R} \to \mathbf{R} : f(x,w) = wm(x,\theta), \theta \in \Theta\}$. If Assumptions 3.2 and 3.4(ii) hold, then $\mathcal{F}$ is Donsker.*

Proof: For any $\theta_1, \theta_2 \in \Theta$, the Cauchy-Schwarz inequality and the mean value theorem imply that

$$\sup_{x\in\mathcal{X}} |w(m^{(i)}(x,\theta_1) - m^{(i)}(x,\theta_2))| \le \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \|\nabla_\theta m(x,\theta)\|_F \times \|\theta_1 - \theta_2\| \times |w| = G(w)\|\theta_1 - \theta_2\| , \tag{B.1}$$

where the equality holds for $G(w) \equiv M|w|$ for some constant M due to Assumption 3.4(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:

$$N_{[\,]}(2\epsilon\|G\|_{L^2}, \mathcal{F}, \|\cdot\|_{L^2}) \le N(\epsilon, \Theta, \|\cdot\|) . \tag{B.2}$$

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

$$\int_0^\infty \sqrt{\log N_{[\,]}(\epsilon, \mathcal{F}, \|\cdot\|_{L^2})}d\epsilon = 2\|G\|_{L^2}\int_0^\infty \sqrt{\log N_{[\,]}(2u\|G\|_{L^2}, \mathcal{F}, \|\cdot\|_{L^2})}du$$
$$\leq 2\|G\|_{L^2}\int_0^\infty \sqrt{N(u,\Theta,\|\cdot\|)}du \leq 2\|G\|_{L^2}\int_0^D \sqrt{d_\theta \log(D/u)}du < \infty \ , \quad \text{(B.3)}$$

where the final inequality holds due to $N(u,\Theta,\|\cdot\|) \leq (\text{diam}(\Theta)/u)^{d_\theta}$. Since $\|G\|_{L^2}^2 = M^2E[W_i^2] < \infty$, the claim of the Lemma then follows from result (B.3) and Theorem 2.5.6 in van der Vaart and Wellner (1996). ■

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

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 $p_f : \mathbf{M}_s \to \mathbf{R}$ 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 $(\mathbf{M}_s, \tau)$ 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 $(\mathbf{M}_s,\tau)$. Hence, letting $N^o(P)$ denote the interior of $N(P)$ (relative to $\mathbf{M}$), we obtain that $N^o(P) = N_s(P) \cap \mathbf{M}$ for some open set $N_s(P) \subseteq \mathbf{M}_s$. However, since $(\mathbf{M}_s,\tau)$ is locally convex, there exists an open (in $\mathbf{M}_s$) convex neighborhood of P with $N_s'(P) \subseteq N_s(P)$. Defining $N'(P) = N_s'(P) \cap \mathbf{M}$ we obtained the desired result by convexity of $\mathbf{M}$. ■

Lemma B.3. *Let Assumptions 3.2, 3.3, 3.4, 3.5 hold and $P \in \mathbf{P}$. For any $Q \in \mathbf{M}$ define $Q_\tau \equiv \tau Q + (1-\tau)P$ and $\Xi(p,Q)$ as in* (A.24)*. Then, there is $N(P) \subseteq \mathbf{M}$ such that for all $(Q,p,\tau_0) \in N(P) \times \mathbb{S}^{d_\theta} \times [0,1]$:*

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

Proof: First observe that by Lemma B.2 we may without loss of generality assume neighborhoods are convex. Hence, if $Q \in N(P)$, then $Q_\tau \in N(P)$ 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 by Lemma A.2 and Assumption 3.5, Lemma A.9 and Corollary 5 in Milgrom and Segal (2002) imply that for Q in a neighborhood of P the function $\tau \mapsto \nu(p,\Theta_0(Q_\tau))$ is directionally differentiable with:

$$\frac{\partial}{\partial\tau_+}\nu(p,\Theta_0(Q_\tau))\Big|_{\tau=\tau_0} = \max_{\theta^*\in\Xi(p,Q_{\tau_0})}\lambda(p,Q_{\tau_0})'\nabla F(\int m(x,\theta^*)dQ_{\tau_0}(x))\int m(x,\theta^*)(dQ(x)-dP(x)) \quad \text{(B.4)}$$
$$\frac{\partial}{\partial\tau_-}\nu(p,\Theta_0(Q_\tau))\Big|_{\tau=\tau_0} = \min_{\theta^*\in\Xi(p,Q_{\tau_0})}\lambda(p,Q_{\tau_0})'\nabla F(\int m(x,\theta^*)dQ_{\tau_0}(x))\int m(x,\theta^*)(dQ(x)-dP(x)) \quad \text{(B.5)}$$

where $\frac{\partial}{\partial\tau_+}$ and $\frac{\partial}{\partial\tau_-}$ denote the right and left derivatives respectively. By Lemma A.10, however, for every $1 \leq i \leq d_F$ such that $\lambda^{(i)}(p,Q_{\tau_0}) \neq 0$ we must have $\theta_1^{(j)} = \theta_2^{(j)}$ for all $j \in \mathcal{S}_i$ and $\theta_1,\theta_2 \in \Xi(p,Q_{\tau_0})$. Therefore, since $A\theta$ does not depend on τ, we immediately can conclude from (3) (4), and results (B.4) and (B.5) that:

$$\frac{\partial}{\partial\tau_+}\nu(p,\Theta_0(Q_\tau))\Big|_{\tau=\tau_0} = \max_{\theta^*\in\Xi(p,Q_{\tau_0})}\sum_{i:\lambda^{(i)}(p,Q_{\tau_0})\neq 0}\lambda^{(i)}(p,Q_{\tau_0})\frac{\partial}{\partial\tau}F_S^{(i)}(\int m_S(x,\theta^*)dQ_\tau(x))\Big|_{\tau=\tau_0}$$
$$= \min_{\theta^*\in\Xi(p,Q_{\tau_0})}\sum_{i:\lambda^{(i)}(p,Q_{\tau_0})\neq 0}\lambda^{(i)}(p,Q_{\tau_0})\frac{\partial}{\partial\tau}F_S^{(i)}(\int m_S(x,\theta^*)dQ_\tau(x))\Big|_{\tau=\tau_0} = \frac{\partial}{\partial\tau_-}\nu(p,\Theta_0(Q_\tau))\Big|_{\tau=\tau_0} . \quad \text{(B.6)}$$

Therefore, we conclude from (B.6) that (B.4) and (B.5) agree, and the Lemma follows. ■

Lemma B.4. *Let $N(P) \subseteq \mathbf{M}$ be a neighborhood of P and $\Gamma : \mathbb{S}^{d_\theta} \times N(P) \to \mathbf{R}^k$ be an upper hemicontinuous correspondence. Then for every $\epsilon > 0$, there exists a $\delta > 0$ and neighborhood $N'(P) \subseteq N(P)$ such that:*

$$\sup_{\|p-\tilde{p}\|<\delta}\sup_{Q\in N'(P)}\sup_{\gamma\in\Gamma(p,Q)}\inf_{\tilde{\gamma}\in\Gamma(\tilde{p},P)}\|\gamma-\tilde{\gamma}\| < \epsilon \ .$$

Proof: Fix $\epsilon > 0$, and for any $\zeta > 0$ and $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P)$ let $\Gamma^\zeta(p, Q) \equiv \{\gamma \in \mathbf{R}^k : \inf_{\tilde{\gamma} \in \Gamma(p,Q)} \|\gamma - \tilde{\gamma}\| < \zeta\}$, and $N_\zeta(p) \equiv \{\tilde{p} \in \mathbb{S}^{d_\theta} : \|p - \tilde{p}\| < \zeta\}$. Since the correspondence $\Gamma : \mathbb{S}^{d_\theta} \times N(P) \to \mathbf{R}^k$ is upper hemicontinuous, for each $p \in \mathbb{S}^{d_\theta}$ there is a $\zeta(p) > 0$ and a neighborhood $N(P|p)$ of P in $\mathbf{M}$ such that:

$$\Gamma(\tilde{p}, Q) \subseteq \Gamma^{\frac{\epsilon}{2}}(p, P) \tag{B.7}$$

for all $(\tilde{p}, Q) \in N_{\zeta(p)}(p) \times N(P|p)$. Since $\{N_{\zeta(p)/2}(p)\}_{p \in \mathbb{S}^{d_\theta}}$ is an open cover of $\mathbb{S}^{d_\theta}$, by compactness, there exists a finite set $\{p_i\}_{i=1}^K$ such that $\{N_{\zeta(p_i)/2}(p_i)\}_{i=1}^K$ is a subcover for $\mathbb{S}^{d_\theta}$. Further let $N'(P) \equiv N(P) \cap \{\bigcap_{i=1}^K N(P|p_i)\}$, and set $\delta \equiv \min_{1 \le i \le K} \zeta(p_i)/2$. Then note that if $p \in N_{\zeta(p_i)/2}(p_i)$ and $\|p - \tilde{p}\| < \delta$, then $p, \tilde{p} \in N_{\zeta(p_i)}(p_i)$. Therefore, since all $p \in \mathbb{S}^{d_\theta}$ satisfy $p \in N_{\zeta(p_i)/2}(p_i)$ for some $1 \le i \le K$ and $N'(P) \subseteq N(P|p_i)$ for all $1 \le i \le K$, we obtain

$$\sup_{\|p-\tilde{p}\|<\delta} \sup_{Q \in N'(P)} \sup_{\gamma \in \Gamma(p,Q)} \inf_{\tilde{\gamma} \in \Gamma(\tilde{p},P)} \|\gamma - \tilde{\gamma}\|$$
$$\le \max_{1 \le i \le K} \sup_{p, \tilde{p} \in N_{\zeta(p_i)}(p_i)} \sup_{Q \in N(P|p_i)} \sup_{\gamma \in \Gamma(p,Q)} \inf_{\tilde{\gamma} \in \Gamma(\tilde{p},P)} \|\gamma - \tilde{\gamma}\| \le \max_{1 \le i \le K} \sup_{\gamma \in \Gamma^{\frac{\epsilon}{2}}(p_i,P)} \inf_{\tilde{\gamma} \in \Gamma(p_i,P)} 2\|\gamma - \tilde{\gamma}\| < \epsilon \ , \tag{B.8}$$

where in the second inequality we employed (B.7) and the third inequality follows by definition of $\Gamma^{\frac{\epsilon}{2}}(p, P)$. ■

Lemma B.5. *Let Assumption 3.1 hold and P_* denote inner probability. Then for every neighborhood $N(P) \subseteq \mathbf{M}$:*

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

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 $p_f : \mathbf{M} \to \mathbf{R}$ 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, $\{U_i\}_{i=1}^K$ open sets in $\mathbf{R}$, and $\{f_i\}_{i=1}^K \in \mathcal{F}$, the sets of the form:

$$\bigcap_{i=1}^K \{Q \in \mathbf{M} : p_{f_i}(Q) \in U_i\} \tag{B.9}$$

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

$$\liminf_{n \to \infty} P_*(\hat{P}_n \in N(P)) \ge \liminf_{n \to \infty} P(\max_{1 \le i \le K_0} |\int f_i(x)(d\hat{P}_n(x) - dP(x))| \le \epsilon) = 1 \ , \tag{B.10}$$

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

Lemma B.6. *If Assumptions 3.2, 3.4(i)-(ii), 3.5 hold and $P \in \mathbf{P}$, then there exists a neighborhood $N(P) \subseteq \mathbf{M}$ of P such that for any $1 \le i \le d_F$ and any $\theta, \tilde{\theta} \in \Theta$ satisfying $\theta^{(j)} = \tilde{\theta}^{(j)}$ for all $j \in \mathcal{S}_i$ it follows that:*

$$\nabla F_S^{(i)}(\int m_S(x,\theta) dQ(x)) m_S(x_0, \theta) = \nabla F_S^{(i)}(\int m_S(x,\tilde{\theta}) dQ(x)) m_S(x_0, \tilde{\theta}) \quad \textit{for all } (Q, x_0) \in N(P) \times \mathcal{X} \ .$$

Proof: By Lemma A.2, there is a neighborhood $N(P) \subseteq \mathbf{M}$ such that the set $\mathcal{R}(Q) \equiv \{\int m(x,\theta) dQ(x)\}_{\theta \in \Theta}$ is compact and satisfies $\mathcal{R}(Q) \subset V_0$ for all $Q \in N(P)$. Letting $\mathcal{R}(Q)^\delta \equiv \{v \in \mathbf{R}^{d_m} : \inf_{\tilde{v} \in \mathcal{R}(Q)} \|v - \tilde{v}\| < \delta\}$, it follows from V_0 being open by Assumption 3.5 that for each $Q \in N(P)$ there exists a $\delta_0(Q) > 0$ such that $\mathcal{R}(Q)^{\delta_0(Q)} \subset V_0$. Moreover, by Assumption 3.4(i), there exists an $M < \infty$ such that $\|m(x,\theta)\| \le M$ for all $(x, \theta) \in \mathcal{X} \times \Theta$. Hence, we obtain that if $c \in \mathbf{R}$ satisfies $|1 - c| < \delta_0(Q)/M$, then $\{c \int m(x,\theta) dQ(x)\}_{\theta \in \Theta} \subseteq \mathcal{R}(Q)^{\delta_0(Q)} \subset V_0$. Therefore, Assumption 3.5(i) implies that for any $Q \in N(P)$, $1 \le i \le d_F$ and $\theta, \tilde{\theta} \in \Theta$ with $\theta^{(j)} = \tilde{\theta}^{(j)}$ for all $j \in \mathcal{S}_i$:

$$\nabla F_S^{(i)}(\int m_S(x,\theta) dQ(x)) \int m_S(x,\theta) dQ(x) = \frac{\partial}{\partial c}\{F_S^{(i)}(c \int m_S(x,\theta) dQ(x))\}\Big|_{c=1}$$
$$= \frac{\partial}{\partial c}\{F_S^{(i)}(c \int m_S(x,\tilde{\theta}) dQ(x))\}\Big|_{c=1} = \nabla F_S^{(i)}(\int m_S(x,\tilde{\theta}) dQ(x)) \int m_S(x,\tilde{\theta}) dQ(x) \ . \tag{B.11}$$

Next, for any $x_0 \in \mathcal{X}$, let $D_{x_0} \in \mathbf{M}$ denote the probability measure satisfying $D_{x_0}(X_i = x_0) = 1$ and define $M_\tau(Q, D_{x_0}) \equiv (1-\tau)Q + \tau D_{x_0}$. Since $M_\tau(Q, D_{x_0}) \to Q$ in the total variation metric as $\tau \to 0$, it follows from $Q \in N(P)$ and $N(P)$ being open, that there is a $\tau_0 > 0$ such that $Q' \equiv M_{\tau_0}(Q, D_{x_0}) \in N(P)$. Thus, Lemma B.2 implies $M_\tau(Q, Q') \in N(P)$ for all $\tau \in [0,1]$, and hence for any $1 \le i \le d_F$ and $\theta, \tilde\theta \in \Theta$ with $\theta^{(j)} = \tilde\theta^{(j)}$ for all $j \in \mathcal{S}_i$

$$\tau_0 \nabla F_S^{(i)}(\int m_S(x,\theta)dQ(x)) \int m_S(x,\theta)(dD_{x_0}(x) - dQ(x)) = \frac{\partial}{\partial\tau}\{F_S^{(i)}(\int m_S(x,\theta)dM_\tau(Q,Q')(x))\}\Big|_{\tau=0}$$
$$= \frac{\partial}{\partial\tau}\{F_S^{(i)}(\int m_S(x,\tilde\theta)dM_\tau(Q,Q')(x))\}\Big|_{\tau=0} = \tau_0 \nabla F_S^{(i)}(\int m_S(x,\tilde\theta)dQ(x)) \int m_S(x,\tilde\theta)(dD_{x_0}(x) - dQ(x)) \ . \quad \text{(B.12)}$$

Therefore, the claim of the Lemma follows from $\tau_0 > 0$ and results (B.11) and (B.12). ■

Lemma B.7. *Let Assumptions 3.2, 3.4, 3.5 and 4.2(i) hold, $P \in \mathbf{P}$, $\Xi(p,P)$ be as in* (A.24) *and $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ satisfy $\theta^*(p) \in \Xi(p,P)$ for all $p \in \mathbb{S}^{d_\theta}$. Then, for each $p \in \mathbb{S}^{d_\theta}$ there exists a map $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ such that*

$$\|\theta^*(p) - \Pi_p\theta\| \le \inf_{\tilde\theta \in \Xi(p,P)} \sqrt{d_\theta}\|\tilde\theta - \theta\| \quad \text{(B.13)}$$

for all $\theta \in \Theta$. In addition, there is a neighborhood $N(P) \subseteq \mathbf{M}$ such that for all $(p, Q, x_0, \theta) \in \mathbb{S}^{d_\theta} \times N(P) \times \mathcal{X} \times \Theta$

$$\lambda(p,P)'\nabla F_S(\int m_S(x,\theta)dQ(x))m_S(x_0,\theta) = \lambda(p,P)'\nabla F_S(\int m_S(x,\Pi_p\theta)dQ(x))m_S(x_0,\Pi_p\theta) \ . \quad \text{(B.14)}$$

Proof: We first construct the map $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$. To this end, for each $p \in \mathbb{S}^{d_\theta}$ we define the set:

$$\mathcal{I}(p) \equiv \bigcup_{i:\lambda^{(i)}(p,P)\neq 0} \mathcal{S}_i \ , \quad \text{(B.15)}$$

and for any $\theta \in \Theta$ let $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ satisfy $(\Pi_p\theta)^{(j)} = \theta^*(p)^{(j)}$ if $j \notin \mathcal{I}(p)$, and $(\Pi_p\theta)^{(j)} = \theta^{(j)}$ if $j \in \mathcal{I}(p)$. Then,

$$\|\theta^*(p) - \Pi_p\theta\| \le \max_{j\in\mathcal{I}(p)} \sqrt{d_\theta}|\theta^*(p)^{(j)} - (\Pi_p\theta)^{(j)}| \le \inf_{\tilde\theta\in\Xi(p,P)} \sqrt{d_\theta}\|\tilde\theta - \theta\| \ , \quad \text{(B.16)}$$

where the first inequality follows from $\theta^*(p)^{(j)} = (\Pi_p\theta)^{(j)}$ for all $j \notin \mathcal{I}(p)$, while the second inequality is the result of $\theta^*(p)^{(j)} = \theta^{(j)}$ for all $\theta \in \Xi(p,P)$ and $j \in \mathcal{I}(p)$ by Lemma A.10, and $\theta^{(j)} = (\Pi_p\theta)^{(j)}$ for all $j \in \mathcal{I}(p)$. Moreover, since for all $1 \le i \le d_F$ such that $\lambda^{(i)}(p,P) \neq 0$ we have $(\Pi_p\theta)^{(j)} = \theta^{(j)}$ for all $j \in \mathcal{S}_i$, it follows from Lemma B.6 that there exists a neighborhood $N(P) \subseteq \mathbf{M}$ such that for all $(p, Q, x_0, \theta) \in \mathbb{S}^{d_\theta} \times N(P) \times \mathcal{X} \times \Theta$

$$\lambda(p,P)'\nabla F_S(\int m_S(x,\theta)dQ(x))m_S(x_0,\theta) = \sum_{i:\lambda^{(i)}(p,P)\neq 0} \lambda^{(i)}(p,P)\nabla F_S^{(i)}(\int m_S(x,\Pi_p\theta)dQ(x))m_S(x_0,\Pi_p\theta)$$
$$= \lambda(p,P)'\nabla F_S(\int m_S(x,\Pi_p\theta)dQ(x))m_S(x_0,\Pi_p\theta) \ . \quad \text{(B.17)}$$

Therefore, the claims of the Lemma follow from results (B.16) and (B.17). ■

Lemma B.8. *Let $\{W_i, X_i\}_{i=1}^n$ be i.i.d. with $W_i \in \mathbf{R}$ independent of X_i and $E[W_i^2] < \infty$. Define $\hat P_{n,\tau} \equiv \tau\hat P_n + (1-\tau)P$ for any $\tau \in [0,1]$ and $\Xi(p,Q)$ as in* (A.24)*. If Assumptions 3.1, 3.2, 3.4, 3.5, 4.2(i) hold, $P \in \mathbf{P}$ and P^W and $\hat P_n^W$ are the population and empirical measures of (X_i, W_i), then uniformly in $(p,\tau) \in \mathbb{S}^{d_\theta} \times [0,1]$ and $\theta \in \Xi(p, \hat P_{n,\tau})$:*

$$\sqrt{n}\lambda(p,P)'\nabla F_S(\int m_S(x,\theta)d\hat P_{n,\tau}(x)) \int w m_S(x,\theta)(d\hat P_n^W(x,w) - dP^W(x,w))$$
$$= \sqrt{n}\lambda(p,P)'\nabla F_S(\int m_S(x,\theta^*(p))dP(x)) \int w m_S(x,\theta^*(p))(d\hat P_n^W(x,w) - dP^W(x,w)) + o_p(1) \quad \text{(B.18)}$$

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

Proof: If $N(P) \subseteq \mathbf{M}$ is convex and $\hat P_n \in N(P)$, then $\hat P_{n,\tau} \in N(P)$ for all $\tau \in [0,1]$. Therefore, by Lemmas A.2, A.7, B.2 and B.5 we obtain that with inner probability tending to one $\{\int m(x,\theta)d\hat P_{n,\tau}\}_{\theta\in\Theta} \subset V_0$ and $\Xi(p, \hat P_{n,\tau})$ is well defined for all $(p,\tau) \in \mathbb{S}^{d_\theta} \times [0,1]$. Next, let $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ be as in Lemma B.7, and note that by (B.13)

$$\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \|\Pi_p\theta - \theta^*(p)\| \leq \sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \inf_{\tilde{\theta}\in\Xi(p,P)} \sqrt{d_\theta}\|\theta - \tilde{\theta}\| = o_p(1) \ , \tag{B.19}$$

where the final result follows from Lemmas A.7, B.2 and B.5, and Lemma B.4 applied with $\Gamma(p,Q) = \Xi(p,Q)$. Moreover, since $\theta^*(p) \in \Theta_0(P)$ for all $p \in \mathbb{S}^{d_\theta}$, results (A.23) and (B.19) further imply that:

$$\liminf_{n\to\infty} P(\Pi_p\theta \in \Theta \text{ for all } \theta \in \Xi(p,\hat{P}_{n,\tau}) \text{ and } (p,\tau) \in \mathbb{S}^{d_\theta} \times [0,1]) = 1 \ . \tag{B.20}$$

Furthermore, by Lemmas B.2, B.5 and B.7, the map $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ satisfies uniformly in $(p,\tau,\theta) \in \mathbb{S}^{d_\theta} \times [0,1] \times \Theta$:

$$\begin{aligned}\sqrt{n}\lambda(p,P)'\nabla F_S(&\int m_S(x,\theta)d\hat{P}_{n,\tau}(x)) \int w m_S(x,\theta)(d\hat{P}_n^W(x,w) - dP^W(x,w)) \\ &= \sqrt{n}\lambda(p,P)'\nabla F_S(\int m_S(x,\Pi_p\theta)d\hat{P}_{n,\tau}(x)) \int w m_S(x,\Pi_p\theta)(d\hat{P}_n^W(x,w) - dP^W(x,w)) + o_p(1) \ . \end{aligned} \tag{B.21}$$

Next, observe that by Lemmas A.2, B.2 and B.5, it follows that for V_0 as in Assumption 3.5 we have:

$$\liminf_{n\to\infty} P(\int m(x,\theta)d\hat{P}_{n,\tau}(x) \in V_0 \text{ for all } (\theta,\tau) \in \Theta \times [0,1]) = 1 \ . \tag{B.22}$$

Assumption 3.2 and (A.3) imply $E[m_S(X_i,\cdot)]$ is uniformly continuous, and hence by (B.19), (B.20) and Lemma B.1:

$$\begin{aligned}\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} &\|\int m_S(x,\Pi_p\theta)d\hat{P}_{n,\tau}(x) - \int m_S(x,\theta^*(p))dP(x)\| \\ &\leq \sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \|\int (m_S(x,\Pi_p\theta) - m_S(x,\theta^*(p)))dP(x)\| + o_p(1) = o_p(1) \ . \end{aligned} \tag{B.23}$$

Thus, ∇F being uniformly continuous on V_0 by Assumption 3.5(ii), (B.20), (B.22), (B.23) and Lemma A.11 imply:

$$\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} |\lambda(p,P)'(\nabla F_S(\int m_S(x,\Pi_p\theta)d\hat{P}_{n,\tau}(x)) - \nabla F_S(\int m_S(x,\theta^*(p))dP(x)))| = o_p(1) \ . \tag{B.24}$$

In addition, also observe that Lemma B.1 allows us to conclude that:

$$\sup_{\theta\in\Theta} \sqrt{n}\|\int w m(x,\theta)(d\hat{P}_n^W(x,w) - dP^W(x,w))\| = O_p(1) \ . \tag{B.25}$$

Therefore, from results (B.20), (B.24) and (B.25) we obtain that uniformly in $(p,\tau) \in \mathbb{S}^{d_\theta} \times [0,1]$ and $\theta \in \Xi(p,\hat{P}_{n,\tau})$:

$$\begin{aligned}\sqrt{n}\lambda(p,P)'\nabla F_S(&\int m_S(x,\Pi_p\theta)d\hat{P}_{n,\tau}(x)) \int w m_S(x,\Pi_p\theta)(d\hat{P}_n^W(x,w) - dP^W(x,w)) \\ &= \sqrt{n}\lambda(p,P)'\nabla F_S(\int m_S(x,\theta^*(p))dP(x)) \int w m_S(x,\Pi_p\theta)(d\hat{P}_n^W(x,w) - dP^W(x,w)) + o_p(1) \ . \end{aligned} \tag{B.26}$$

To conclude, we note that (B.19), (B.20) and Lemma B.1 imply that for some deterministic sequence $\delta_n \downarrow 0$,

$$\begin{aligned}\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} &\sqrt{n}\|\int w(m_S(x,\Pi_p\theta) - m_S(x,\theta^*(p)))(d\hat{P}_n^W(x,w) - dP^W(x,w))\| \\ &\leq \sup_{\|\theta_1-\theta_2\|<\delta_n} \sqrt{n}\|\int w(m_S(x,\theta_1) - m_S(x,\theta_2))(d\hat{P}_n^W(x,w) - dP^W(x,w))\| + o_p(1) = o_p(1) \ . \end{aligned} \tag{B.27}$$

Moreover, note that since $P \in \mathbf{P}$ satisfies Assumption 3.6(iii), it follows from Lemma A.11 and Assumption 3.5(ii) that $\|\lambda(p,P)'\nabla F_S(\int m_S(x,\theta^*(p))dP(x))\|$ is uniformly bounded in $p \in \mathbb{S}^{d_\theta}$. Hence, by (B.27) and Cauchy-Schwarz,

$$\begin{aligned}\sqrt{n}\lambda(p,P)'\nabla F_S(&\int m_S(x,\theta^*(p))dP(x)) \int w m_S(x,\Pi_p\theta)(d\hat{P}_n^W(x,w) - dP^W(x,w)) \\ &= \sqrt{n}\lambda(p,P)'\nabla F_S(\int m_S(x,\theta^*(p))dP(x)) \int w m_S(x,\theta^*(p))(d\hat{P}_n^W(x,w) - dP^W(x,w)) \ , \end{aligned} \tag{B.28}$$

uniformly in $(p,\tau) \in \mathbb{S}^{d_\theta} \times [0,1]$ and $\theta \in \Xi(p, \hat{P}_{n,\tau})$. The Lemma then follows from (B.21), (B.26) and (B.28). ■

Lemma B.9. *Let Assumptions 3.1, 3.2, 3.3, 3.4, 3.5 hold, $P \in \mathbf{P}$ and $\Xi(p,P)$ be as in* (A.24)*. Then:*

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

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

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

$$\begin{aligned}&\sqrt{n}\{\nu(p,\Theta_0(\hat{P}_n)) - \nu(p,\Theta_0(P))\} \\ &\qquad = \sqrt{n}\lambda(p,\hat{P}_{n,\tau_0(p)})'\nabla F(\int m(x,\tilde{\theta}(p))d\hat{P}_{n,\tau_0(p)}(x))\int m(x,\tilde{\theta}(p))(d\hat{P}_n(x) - dP(x)) + o_p(1)\end{aligned} \quad \text{(B.29)}$$

for some $\tau_0 : \mathbb{S}^{d_\theta} \to (0,1)$ and $\tilde{\theta} : \mathbb{S}^{d_\theta} \to \Theta$ such that $\tilde{\theta}(p) \in \Xi(p,\hat{P}_{n,\tau_0(p)})$ for all $p \in \mathbb{S}^{d_\theta}$. Next, fix $\epsilon > 0$ and note that by Lemmas A.9 and A.12 there exists a neighborhood $N(P) \subseteq \mathbf{M}$ such that the correspondence $(p,Q) \mapsto \lambda(p,Q)$ is upper hemicontinuous and singleton valued for all $(p,Q) \in \mathbb{S}^{d_\theta} \times N(P)$. Applying Lemmas B.2 and B.4 with $\Gamma(p,Q) = \lambda(p,Q)$ then implies that there exists a convex neighborhood $N'(P) \subseteq N(P) \subseteq \mathbf{M}$ such that:

$$\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{Q\in N'(P)} \|\lambda(p,Q) - \lambda(p,P)\| < \epsilon \ . \quad \text{(B.30)}$$

Since $N'(P)$ is convex, $\hat{P}_n \in N'(P)$ implies $\hat{P}_{n,\tau} \in N'(P)$ for all $\tau \in [0,1]$. Therefore, we are able to conclude that:

$$\liminf_{n\to\infty} P(\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\tau\in[0,1]} \|\lambda(p,\hat{P}_{n,\tau}) - \lambda(p,P)\| < \epsilon) \geq \liminf_{n\to\infty} P(\hat{P}_n \in N'(P)) = 1 \ , \quad \text{(B.31)}$$

where the final equality follows from Lemma B.5. Thus, result (B.22) and Assumption 3.5(ii), result (B.25) applied with the random variable $W_i = 1$ almost surely, and results (B.29) and (B.31) in turn imply uniformly in $p \in \mathbb{S}^{d_\theta}$:

$$\begin{aligned}&\sqrt{n}\{\nu(p,\Theta_0(\hat{P}_n)) - \nu(p,\Theta_0(P))\} \\ &\qquad = \sqrt{n}\lambda(p,P)'\nabla F(\int m(x,\tilde{\theta}(p))d\hat{P}_{n,\tau_0(p)}(x))\int m(x,\tilde{\theta}(p))(d\hat{P}_n(x) - dP(x)) + o_p(1) \\ &\qquad = \sqrt{n}\lambda(p,P)'\nabla F(\int m(x,\theta^*(p))dP(x))\int m(x,\theta^*(p))(d\hat{P}_n(x) - dP(x)) + o_p(1) \ ,\end{aligned} \quad \text{(B.32)}$$

where the second equality follows from (3) and $\int A\theta(d\hat{P}_n(x) - dP(x)) = 0$ for all $\theta \in \Theta$, and Lemma B.8 applied with the random variable $W_i = 1$ almost surely. ■

Lemma B.10. *Let Assumptions 3.1, 3.2, 3.3, 3.4, 3.5 hold, $P \in \mathbf{P}$, $\Xi(p,P)$ be as in* (A.24) *and $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ satisfy $\theta^*(p) \in \Xi(p,P)$ for all $p \in \mathbb{S}^{d_\theta}$. Then the following class is Donsker in $\mathcal{C}(\mathbb{S}^{d_\theta})$:*

$$\mathcal{F} \equiv \{f : \mathcal{X} \to \mathbf{R} : f(x) = \lambda(p,P)'H(\theta^*(p))m(x,\theta^*(p)) \text{ for some } p \in \mathbb{S}^{d_\theta}\} \ .$$

Proof: For notational simplicity, let $H_S(\theta) \equiv \nabla F_S(\int m_S(x,\theta)dP(x))$, $H_S^{(i)}(\theta) \equiv \nabla F_S^{(i)}(\int m_S(x,\theta)dP(x))$, and:

$$G_n(p) \equiv \sqrt{n}\lambda(p,P)'H(\theta^*(p))\int m(x,\theta^*(p))(d\hat{P}_n(x) - dP(x)) \ . \quad \text{(B.33)}$$

We first note that since $\lambda(\cdot,P)$, m and $H(\cdot)$ are bounded by Lemma A.11, Assumption 3.4(i), Assumption 3.5(ii) and $P \in \mathbf{P}$ satisfying Assumption 3.6(iii), it follows from the central limit theorem that for any $p \in \mathbb{S}^{d_\theta}$

$$G_n(p) \xrightarrow{L} N(0,\sigma^2(p)) \ , \quad \text{(B.34)}$$

where $\sigma^2(p) \equiv \text{Var}(\lambda(p,P)'H(\theta^*(p))m(X_i,\theta^*(p)))$. Moreover, also observe that since $\int A\theta(d\hat{P}_n(x) - dP(x)) = 0$,

$$\begin{aligned} G_n(p) &= \sqrt{n}\lambda(p,P)'H_S(\theta^*(p)) \int m_S(x,\theta^*(p))(d\hat{P}_n(x) - dP(x)) \\ &= \sqrt{n} \sum_{i:\lambda^{(i)}(p,P)\neq 0} \lambda^{(i)}(p,P)H_S^{(i)}(\theta^*(p)) \int m_S(x,\theta^*(p))(d\hat{P}_n(x) - dP(x)) \ . \end{aligned} \tag{B.35}$$

Thus, result (B.35), and Lemmas A.10 and B.6 imply $G_n(p)$ is independent of how $\theta^*(p) \in \Xi(p,P)$ is selected, and hence so is the asymptotic variance $\sigma^2(p)$.

Note that in (B.34) it was argued that $G_n(p)$ is bounded in $p \in \mathbb{S}^{d_\theta}$, while identical arguments to those in (A.50)-(A.51) 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. Equivalently, we aim to show

$$\sup_{\|p-\tilde{p}\|<\delta_n} |G_n(p) - G_n(\tilde{p})| = o_p(1) \ , \tag{B.36}$$

for any sequence $\delta_n \downarrow 0$. First observe that compactness of $\mathbb{S}^{d_\theta}$ and Lemma A.12 imply $\lambda(\cdot,P) : \mathbb{S}^{d_\theta} \to \mathbf{R}^{d_F}$ is uniformly continuous. Therefore, by Assumption 3.5(ii), $P \in \mathbf{P}$ satisfying Assumption 3.6(iii) and result (B.25):

$$\begin{aligned} &\sup_{\|p-\tilde{p}\|<\delta_n} \sqrt{n}|(\lambda(p,P) - \lambda(\tilde{p},P))'H_S(\theta^*(\tilde{p})) \int m_S(x,\theta^*(\tilde{p}))(d\hat{P}_n(x) - dP(x))| \\ &\quad \le \sup_{\|p-\tilde{p}\|<\delta_n} \|\lambda(p,P) - \lambda(\tilde{p},P)\| \times \sup_{v\in V_0} \|\nabla F(v)\|_F \times \sup_{\theta\in\Theta} \|\sqrt{n} \int m(x,\theta)(d\hat{P}_n(x) - dP(x))\| = o_p(1) \ . \end{aligned} \tag{B.37}$$

Hence, by results (B.35) and (B.37) we obtain by Lemma B.7 that for some mapping $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ satisfying (B.14):

$$\begin{aligned} &\sup_{\|p-\tilde{p}\|<\delta_n} |G_n(p) - G_n(\tilde{p})| \\ &\le \sup_{\|p-\tilde{p}\|<\delta_n} \sqrt{n}|\lambda(p,P)' \int (H_S(\theta^*(p))m_S(x,\theta^*(p)) - H_S(\theta^*(\tilde{p}))m_S(x,\theta^*(\tilde{p})))(d\hat{P}_n(x) - dP(x))| + o_p(1) \\ &= \sup_{\|p-\tilde{p}\|<\delta_n} \sqrt{n}|\lambda(p,P)' \int (H_S(\theta^*(p))m_S(x,\theta^*(p)) - H_S(\Pi_p\theta^*(\tilde{p}))m_S(x,\Pi_p\theta^*(\tilde{p})))(d\hat{P}_n(x) - dP(x))| + o_p(1) \end{aligned} \tag{B.38}$$

Moreover, it also follows from $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ satisfying condition (B.13), and Lemmas A.7 and B.4 that:

$$\sup_{\|p-\tilde{p}\|<\delta_n} \|\theta^*(p) - \Pi_p\theta^*(\tilde{p})\| \le \sup_{\|p-\tilde{p}\|<\delta_n} \sup_{\tilde{\theta}\in\Xi(\tilde{p},P)} \inf_{\theta\in\Xi(p,P)} \sqrt{d_\theta}\|\theta - \tilde{\theta}\| = o(1) \ . \tag{B.39}$$

Therefore, results (A.23) and (B.39) imply that for δ_n sufficiently small, $\Pi_p\theta^*(\tilde{p}) \in \Theta$ for all $\tilde{p}, p \in \mathbb{S}^{d_\theta}$ with $\|\tilde{p} - p\| < \delta_n$. Hence, from (B.38) and (B.39) we conclude that for some sequence $\gamma_n \to 0$ depending on δ_n,

$$\begin{aligned} &\sup_{\|p-\tilde{p}\|<\delta_n} |G_n(p) - G_n(\tilde{p})| \\ &\qquad \le \sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\|\theta-\tilde{\theta}\|<\gamma_n} \sqrt{n}|\lambda(p,P)' \int (H_S(\theta)m_S(x,\theta) - H_S(\tilde{\theta})m_S(x,\tilde{\theta}))(d\hat{P}_n(x) - dP(x))| + o_p(1) \ , \end{aligned} \tag{B.40}$$

where $\theta, \tilde{\theta}$ are restricted to lie in Θ. However, note $\int m(x,\cdot)dP(x) : \Theta \to \mathbf{R}^{d_m}$ is uniformly continuous by (A.3) and Assumption 3.2(i), and therefore Assumption 3.5(ii) and $P \in \mathbf{P}$ satisfying Assumption 3.6(iii) imply $\theta \mapsto H_S(\theta)$ is uniformly continuous. Therefore, $\lambda(\cdot,P)$ being bounded by Lemma A.11 and result (B.25) imply:

$$\begin{aligned} &\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\|\theta-\tilde{\theta}\|<\gamma_n} \sqrt{n}|\lambda(p,P)'(H_S(\theta) - H_S(\tilde{\theta})) \int m_S(x,\tilde{\theta})(d\hat{P}_n(x) - dP(x))| \\ &\quad \le \sup_{p\in\mathbb{S}^{d_\theta}} \|\lambda(p,P)\| \times \sup_{\|\theta-\tilde{\theta}\|<\gamma_n} \|H_S(\theta) - H_S(\tilde{\theta})\|_F \times \sup_{\theta\in\Theta} \|\sqrt{n} \int m(x,\theta)(d\hat{P}_n(x) - dP(x))\| = o_p(1) \ . \end{aligned} \tag{B.41}$$

In turn, it also follows from $H_S(\theta)$ being uniformly bounded in $\theta \in \Theta$ due to it being continuous and Assumption 3.2(i), Lemma A.11 implying $\|\lambda(p, P)\|$ is uniformly bounded in $p \in \mathbb{S}^{d_\theta}$ and Lemma B.1 that:

$$\sup_{p\in\mathbb{S}^{d_\theta}} \sup_{\|\theta-\tilde{\theta}\|<\gamma_n} \sqrt{n}|\lambda(p, P)' H_S(\theta) \int (m_S(x, \theta) - m_S(x, \tilde{\theta}))(d\hat{P}_n(x) - dP(x))|$$

$$\leq \sup_{p\in\mathbb{S}^{d_\theta}} \|\lambda(p, P)\| \times \sup_{\theta\in\Theta} \|H_S(\theta)\|_F \times \sup_{\|\theta-\tilde{\theta}\|<\gamma_n} \|\sqrt{n} \int (m_S(x, \theta) - m_S(x, \tilde{\theta}))(d\hat{P}_n(x) - dP(x))\| = o_p(1) \quad \text{(B.42)}$$

Hence, we conclude from (B.40), (B.41) and (B.42) that (B.36) holds, which establishes the asymptotic uniform equicontinuity of G_n. In turn, because $\mathbb{S}^{d_\theta}$ 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 B.9, $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ has an influence function $\psi : \mathcal{X} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ given by:

$$\psi(x) \equiv \lambda(\cdot, P)' H(\theta^*(\cdot))\{m(x, \theta^*(\cdot)) - E[m(X_i, \theta^*(\cdot))]\} \quad \text{(B.43)}$$

where $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ with $\theta^*(p) \in \Xi(p, P)$, 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 B.10 and Theorem 18.1 in Kosorok (2008), which establishes (i). The stated convergence in distribution is then immediate from Lemmas B.9 and B.10, 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: Since $L : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}_+$ is a subconvex function and $\{T_n\}$ is a regular estimator, we obtain from Theorems A.1, A.2 and Proposition 5.2.1 in Bickel et al. (1993) that:

$$\liminf_{n\to\infty} E[L(\sqrt{n}\{T_n - \nu(\cdot, \Theta_0(P))\})] \geq E[L(\mathbb{G}_0)] \ . \quad \text{(B.44)}$$

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

$$G_n(p) \equiv \sqrt{n}\{\nu(p, \Theta_0(\hat{P}_n)) - \nu(p, \Theta_0(P))\} \ , \quad \text{(B.45)}$$

and note $G_n \in \mathcal{C}(\mathbb{S}^{d_\theta})$ almost surely. Since L is continuous on $\mathbb{D}_0 \subseteq \mathcal{C}(\mathbb{S}^{d_\theta})$ and $P(\mathbb{G}_0 \in \mathbb{D}_0) = 1$, Theorem 4.1 and Theorem 1.3.6 in van der Vaart and Wellner (1996) imply $L(G_n) \xrightarrow{L} L(\mathbb{G}_0)$ (in $\mathbf{R}$). Hence, since $a \mapsto a \wedge C$ is continuous and bounded on $\mathbf{R}$ for any constant $C > 0$, the Portmanteau Theorem yields:

$$\limsup_{C\uparrow\infty} \limsup_{n\to\infty} |E[L(G_n) \wedge C] - E[L(\mathbb{G}_0) \wedge C]| = 0 \ . \quad \text{(B.46)}$$

Moreover, $L(\mathbb{G}_0) \leq M_0 + M_1\|\mathbb{G}_0\|_\infty^\kappa$ by hypothesis, and therefore Proposition A.2.3 in van der Vaart and Wellner (1996) yields $E[L(\mathbb{G}_0)] \leq M_0 + M_1 E[\|\mathbb{G}_0\|_\infty^\kappa] < \infty$. Therefore, by the monotone convergence theorem:

$$\limsup_{C\uparrow\infty} |E[L(\mathbb{G}_0)] - E[L(\mathbb{G}_0) \wedge C]| = 0 \ . \quad \text{(B.47)}$$

By Assumption 3.5(ii) and Lemmas A.2, A.11 and B.2 there exists a convex neighborhood $N(P) \subseteq \mathbf{M}$ such that: (i) $\nabla F(\int m(x, \theta)dQ(x))$ is uniformly bounded in $(\theta, Q) \in \Theta \times N(P)$; (ii) $\lambda(p, Q)$ is uniformly bounded on $(p, Q) \in \mathbb{S}^{d_\theta} \times N(P)$; and (iii) the conditions of Lemma B.3 are satisfied for all $Q \in N(P)$. For every $\tau \in [0, 1]$ define $\hat{P}_{n,\tau} \equiv \tau\hat{P}_n + (1 - \tau)P$ and note that if $\hat{P}_n \in N(P)$ then (B.29) holds so that uniformly in $p \in \mathbb{S}^{d_\theta}$:

$$G_n = \tilde{\Delta}_n \qquad \tilde{\Delta}_n(p) \equiv \lambda(p, \hat{P}_{n,\tau_0(p)})' \nabla F(\int m(x, \tilde{\theta}(p))d\hat{P}_{n,\tau_0(p)}(x)) \int \sqrt{n} m(x, \tilde{\theta}(p))(d\hat{P}_n(x) - dP(x)) \quad \text{(B.48)}$$

for some $\tau_0 : \mathbb{S}^{d_\theta} \to (0, 1)$ and $\tilde{\theta} : \mathbb{S}^{d_\theta} \to \Theta$ with $\tilde{\theta}(p) \in \Xi(p, \hat{P}_{n,\tau_0(p)})$ for $\Xi(p, Q)$ as in (A.24) (and set $\tilde{\Delta}_n = 0$ if $\hat{P}_n \notin N(P)$). By compactness of Θ, definition of $N(P)$ and m being bounded by Assumption 3.4(i), we must have

$$\max\{\|G_n\|_\infty, \|\tilde{\Delta}_n\|_\infty\} \leq \sqrt{n}C_0 \ , \quad \text{(B.49)}$$

for some $C_0 > 0$. Therefore, $L(f) \leq M_0 + M_1\|f\|_\infty^\kappa$ for all $f \in \mathcal{C}(\mathbb{S}^{d_\theta})$, (B.48) holding if $\hat{P}_n \in N(P)$ and (B.49) yield:

$$\limsup_{n\to\infty} |E[L(G_n)] - E[L(\tilde{\Delta}_n)]| \leq \limsup_{n\to\infty} 2(M_0 + M_1 C_0^\kappa n^{\frac{\kappa}{2}}) P(\hat{P}_n \notin N(P)) \ . \tag{B.50}$$

However, as shown in (B.10), there exists a finite collection $\{f_j\}_{j=1}^{K_0}$ of bounded functions and an $\epsilon > 0$ such that $\{Q \in \mathbf{M} : \max_{1\leq j\leq K_0} |\int f_j(x)(dQ(x) - dP(x))| \leq \epsilon\} \subseteq N(P)$. Therefore, (B.50) and Bernstein's inequality imply:

$$\limsup_{n\to\infty} |E[L(G_n)] - E[L(\tilde{\Delta}_n)]| \leq 2(M_0 + M_1 C_0^\kappa) \limsup_{n\to\infty} \sum_{j=1}^{K_0} n^{\frac{\kappa}{2}} P(|\int f_j(x)(d\hat{P}_n(x) - dP(x))| > \epsilon) = 0 \ . \tag{B.51}$$

From result (B.51) and applying Cauchy-Schwarz and Markov's inequalities we can then conclude that:

$$\limsup_{n\to\infty} |E[L(G_n)] - E[L(G_n) \wedge C]| = \limsup_{n\to\infty} |E[L(\tilde{\Delta}_n)] - E[L(\tilde{\Delta}_n) \wedge C]|$$
$$\leq \limsup_{n\to\infty} E[L(\tilde{\Delta}_n) 1\{L(\tilde{\Delta}_n) > C\}] \leq \limsup_{n\to\infty} \frac{1}{C} E[L^2(\tilde{\Delta}_n)] \ . \tag{B.52}$$

By construction of $N(P)$, 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 $(p, \theta, Q) \in \mathbb{S}^{d_\theta} \times \Theta \times N(P)$. Let $\mathcal{G} \equiv \{g : \mathcal{X} \to \mathbf{R} : g(x) = c'm(x,\theta) \text{ for some } (c,\theta) \in \mathbf{C} \times \Theta\}$ and note that by Assumption 3.4(i) 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 Assumptions 3.4(i)-(ii) that:

$$\sup_{x\in\mathcal{X}} |c_1' m(x,\theta_1) - c_2' m(x,\theta_2)|$$
$$\leq \{\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \|m(x,\theta)\| + \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \|\nabla_\theta m(x,\theta)\|_F \times \sup_{c\in\mathbf{C}} \|c\|\} \times \{\|c_1 - c_2\| + \|\theta_1 - \theta_2\|\} \ , \tag{B.53}$$

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:

$$N_{[\,]}(2\epsilon C_1, \mathcal{G}, \|\cdot\|_\infty) \leq N(\epsilon, \Theta \times \mathbf{C}, \|\cdot\| + \|\cdot\|) \lesssim \epsilon^{-(d_m + d_\theta)} \ . \tag{B.54}$$

Consequently, since $\tilde{\Delta}_n = 0$ whenever $\hat{P}_n \notin N(P)$, the inequality $L(f) \leq M_0 + M_1\|f\|_\infty^\kappa$ for all $f \in \mathcal{C}(\mathbb{S}^{d_\theta})$ implies:

$$\limsup_{n\to\infty} E[L^2(\tilde{\Delta}_n)] \leq \limsup_{n\to\infty} \{2M_0^2 + 2M_1^2 E[\|\tilde{\Delta}_n\|_\infty^{2\kappa}]\}$$
$$\leq \limsup_{n\to\infty} \{2M_0^2 + 2M_1^2 E[\sup_{g\in\mathcal{G}} |\frac{1}{\sqrt{n}} \sum_{i=1}^n \{g(X_i) - E[g(X_i)]\}|^{2\kappa}]\}$$
$$\lesssim 2M_0^2 + (\int_0^1 \sqrt{1 + \log N_{[\,]}(\epsilon C_1, \mathcal{G}, \|\cdot\|_\infty)} d\epsilon)^{2\kappa} \tag{B.55}$$

where the third inequality follows from Theorem 2.14.1 in van der Vaart and Wellner (1996). Combining results (B.52), (B.54) and (B.55), we can finally obtain:

$$\limsup_{C\uparrow\infty} \limsup_{n\to\infty} |E[L(G_n)] - E[L(G_n) \wedge C]| \leq \limsup_{C\uparrow\infty} \limsup_{n\to\infty} \frac{1}{C} E[L^2(\Delta_n)] = 0 \ . \tag{B.56}$$

The claim of the Theorem then follows from results (B.46), (B.47) and (B.56). ■

Proof of Corollary 4.1: For any convex compact valued set K_n, Corollary 1.10 in Li et al. (2002) implies that

$$\sqrt{n} d_H(K_n, \Theta_0(P)) = \sqrt{n}\|\nu(\cdot, K_n) - \nu(\cdot, \Theta_0(P))\|_\infty \ , \tag{B.57}$$

and in particular $\sqrt{n} d_H(\hat{\Theta}_n, \Theta_0(P)) = \sqrt{n}\|\nu(\cdot, \Theta_0(\hat{P}_n)) - \nu(\cdot, \Theta_0(P))\|_\infty$. Therefore, the claim of the Corollary follows if we can verify the conditions of Theorem 4.2 under the loss function $\bar{L} : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}_+$ given by $\bar{L}(f) = L(\|f\|_\infty)$. To this end, note $\bar{L}(f) = L(\|f\|_\infty) = L(\|-f\|_\infty) = \bar{L}(-f)$. Moreover, since $L : \mathbf{R}_+ \to \mathbf{R}_+$ is subconvex, it follows that $0 = L(0) \leq L(a)$, and hence if $L(a) = c$ then by convexity of $\{a : L(a) \leq c\}$ we must have $L(\lambda a) \leq c$ for all

$\lambda \in [0,1]$. In particular, it follows that $L : \mathbf{R}_+ \to \mathbf{R}_+$ is nondecreasing. Therefore, if $\bar{L}(f_1) \leq c$ and $\bar{L}(f_2) \leq c$, then

$$\bar{L}(\lambda f_1 + (1-\lambda) f_2) = L(\|\lambda f_1 + (1-\lambda) f_2\|_\infty) \leq L(\lambda \|f_1\|_\infty + (1-\lambda)\|f_2\|_\infty) \leq c \ , \tag{B.58}$$

where the first inequality follows from L being nondecreasing, and the second by subconvexity of L. It follows from (B.58) that $\bar{L} : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}_+$ is subconvex. The other conditions on $\bar{L}$ have been directly assumed, and the claim of the Corollary follows from Theorem 4.2. ■

Appendix C - Proof of Theorem 4.3

The proof of Theorem 4.3 proceeds by: (i) Deriving the semiparametric efficiency bound, and (ii) Establishing $\{\nu_{|\mathbb{C}}(\cdot, \Theta_0(\hat{P}_n))\}$ attains the bound. The efficiency bound is derived in Theorem C.1, after verifying $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$ is weak-pathwise differentiable (Lemma C.4) and characterizing the tangent space (Lemma C.3). A key challenge in the latter is showing P satisfying Assumption 4.1 does not affect the tangent space (Lemma C.2). The fact that $\{\nu_{|\mathbb{C}}(\cdot, \Theta_0(\hat{P}_n))\}$ attains the efficiency bound follows readily after characterizing its influence function (Lemma C.6).

Some of the derivations in this Appendix are similar to those in Appendices A and B. For conciseness, we provide more succinct derivations but include references to previous instances where analogous arguments were employed.

Lemma C.1. *Let $\mathbf{S_L} \equiv \{s \in L^2_\mu : s = \sqrt{dP/d\mu}$ for some $P \in \mathbf{P_L}\}$, and Assumptions 3.2, 3.4, 3.5 and 4.2(i) hold. If $\eta \mapsto h_\eta$ is a curve in $\mathbf{S_L}$, then there is a neighborhood $N \subseteq \mathbf{R}$ of 0 such that for all $(p, \eta_0) \in \mathbb{C} \times N$, $(p,\eta_0) \mapsto \frac{\partial}{\partial \eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_0}$ exists, satisfies* (A.42) *and is both bounded and continuous on $\mathbb{C} \times N$.*

Proof: First note $\mathbf{P_L} \subseteq \mathbf{P}$ implies $\mathbf{S_L} \subseteq \mathbf{S}$. Therefore, there is a neighborhood $N_1 \subseteq \mathbf{R}$ of 0 such that (A.45) and (A.46) hold for all $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N_1$. Since for any $(p, \eta_0) \in \mathbb{C} \times N_1$, $\Xi(p, P_{\eta_0})$ is a singleton due to $P_{\eta_0} \in \mathbf{P_L}$, it follows that (A.45) and (A.46) equal each other and hence $\frac{\partial}{\partial \eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_0}$ exists and is given by (A.42) for all $(p, \eta_0) \in \mathbb{C} \times N_1$. The existence of a neighborhood $N_2 \subseteq N_1$ such that $(p, \eta_0) \mapsto \frac{\partial}{\partial \eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_0}$ is uniformly bounded in $(p, \eta_0) \in \mathbb{C} \times N_2$ then follows from (A.48), Lemmas A.2 and A.11, and Assumptions 3.4(i) and 3.5(ii).

In order to establish continuity, note that Lemmas A.7 and A.12 imply there is a neighborhood $N \subseteq N_2 \subseteq \mathbf{R}$ such that $(p, \eta_0) \mapsto \lambda(p, P_{\eta_0})$ and $(p, \eta_0) \mapsto \Xi(p, P_{\eta_0})$ are continuous and upper hemicontinuous respectively on $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N$. Next, let $(p_0, \eta_0) \in \mathbb{C} \times N$ and $\{(p_n, \eta_n)\}_{n=1}^\infty$ be a sequence such that $(p_n, \eta_n) \to (p_0, \eta_0)$ and $(p_n, \eta_n) \in \mathbb{C} \times N$ for all n. Since $(p_n, P_{\eta_n}) \in \mathbb{C} \times \mathbf{P_L}$ for all $0 \leq n < \infty$, $\Xi(p_n, P_{\eta_n}) = \{\theta_n^*\}$ for some $\theta_n^* \in \Theta$ and by upper hemicontinuity $\theta_n^* \to \theta_0^*$ with $\Xi(p_0, P_{\eta_0}) = \{\theta_0^*\}$. Result (A.50) and continuity of $(p, P) \mapsto \lambda(p, P)$ then imply:

$$\lim_{n\to\infty} \frac{\partial}{\partial \eta}\nu(p_n, \Theta_0(P_\eta))\Big|_{\eta=\eta_n} = \frac{\partial}{\partial \eta}\nu(p_0, \Theta_0(P_\eta))\Big|_{\eta=\eta_0} \tag{C.1}$$

due to $\frac{\partial}{\partial \eta}\nu(p, \Theta_0(P_\eta))|_{\eta=\eta_n}$ satisfying (A.42) for all integer $0 \leq n < \infty$. ■

Lemma C.2. *If Assumptions 3.2, 3.4, 3.5, 4.2 hold and $\mathbb{C}$ is compact, then the following set is open in* $\mathbf{M}$*:*

$$\mathbf{M_L} \equiv \{P \in \mathbf{M} : \textit{Assumptions 3.6(i)-(iv) and 4.1(i) hold}\} \ . \tag{C.2}$$

Proof: The proof is by contradiction. Suppose there exists a $P \in \mathbf{M_L}$ such that $N(P) \subsetneq \mathbf{M_L}$ for all neighborhoods $N(P) \subseteq \mathbf{M}$ of P. Let $\mathfrak{N}_P$ be the neighborhood system of P with direction $V \succeq W$ whenever $V \subseteq W$, and recall that Lemmas A.2 and A.8, Corollary A.3, and $P \in \mathbf{M_L}$ satisfying Assumption 3.6(ii) imply the set of $P \in \mathbf{M}$ satisfying Assumptions 3.6(i)-(iv) is open in $\mathbf{M}$. Therefore, if the Lemma is false, then for $\mathfrak{A} = \mathfrak{N}_P$ there is a net $\{Q_\alpha\}_{\alpha \in \mathfrak{A}}$ with $Q_\alpha \to P$ such that for each $\alpha \in \mathfrak{A}$: (i) Q_α satisfies Assumption 3.6(i)-(iv), and (ii) there is a $p_\alpha \in \mathbb{C}$ with $\Xi(p_\alpha, Q_\alpha)$ (as in (A.24)) not a singleton. Furthermore, by arguing as in (A.13)-(A.15), there is a $\theta_\alpha \in \Xi(p_\alpha, Q_\alpha)$ with:

$$\mathcal{A}(\theta_\alpha, Q_\alpha) = \bigcap_{\theta \in \Xi(p_\alpha, Q_\alpha)} \mathcal{A}(\theta, Q_\alpha) \ . \tag{C.3}$$

By compactness of $\mathbb{C}$, finiteness of the number of constraints, and Lemma A.7, we can then pass to a subnet $\{Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}\}_{\beta\in\mathfrak{B}}$ such that for some $(p^*, \theta^*) \in \mathbb{C} \times \Xi(p^*, P)$ and a fixed set $\mathcal{C} \subseteq \{1, \ldots, d_F\}$:

$$(Q_{\alpha_\beta}, p_{\alpha_\beta}, \theta_{\alpha_\beta}) \to (P, p^*, \theta^*) \qquad \text{and} \qquad \mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) = \mathcal{C} \ \ \forall \beta \in \mathfrak{B} \ . \tag{C.4}$$

Next, note Assumption 4.2(ii) implies we can partition $\{1, \ldots, d_F\}$ into $\mathcal{I}_L \equiv \{i : \mathcal{S}_i = \emptyset\}$ and $\mathcal{I}_S \equiv \{i : \mathcal{S}_i = \{1, \ldots, d_\theta\}\}$. Since Assumption 3.2(i) and Q_{α_β} satisfying Assumption 3.6(ii) imply $\Xi(p_{\alpha_\beta}, Q_{\alpha_\beta})$ is convex and $F^{(i)}(\int m(x,\cdot)dQ_{\alpha_\beta}(x)) : \Theta \to \mathbf{R}$ is strictly convex for all $i \in \mathcal{I}_S$, $\Xi(p_{\alpha_\beta}, Q_{\alpha_\beta})$ being nonsingleton and (C.3) yield

$$\mathcal{C} \subseteq \mathcal{I}_L \ . \tag{C.5}$$

Hence, by the complementary slackness condition $\lambda^{(i)}(p_{\alpha_\beta}, Q_{\alpha_\beta}) = 0$ for all $i \in \mathcal{I}_S$. Since Theorem 8.3.1 in Luenberger (1969) implies θ_{α_β} is a maximizer of (A.29), we obtain from the first order conditions and $\mathcal{S}_i = \emptyset$ for all $i \in \mathcal{I}_L$:

$$F_A(\int m_A(x)dQ_{\alpha_\beta}(x))'\lambda(p_{\alpha_\beta}, Q_{\alpha_\beta}) = -p_{\alpha_\beta} \ , \tag{C.6}$$

where we exploited $\theta_{\alpha_\beta} \in \Theta^o$ due to Q_{α_β} satisfying Assumption 3.6(i). Since by construction, $\mathcal{A}(\theta_{\alpha_\beta}, Q_{\alpha_\beta}) = \mathcal{C}$, we may let $\lambda^{\mathcal{C}}(p_{\alpha_\beta}, Q_{\alpha_\beta})$, $F_A^{\mathcal{C}}(\int m_A(x)dQ_{\alpha_\beta}(x))$ and $F_S^{\mathcal{C}}(\int m_S(x,\theta)dQ_{\alpha_\beta}(x))$ respectively be the $\#\mathcal{C} \times 1$ subvector of $\lambda(p_{\alpha_\beta}, Q_{\alpha_\beta})$, $\#\mathcal{C} \times d_\theta$ submatrix of $F_A(\int m_A(x)dQ_{\alpha_\beta}(x))$ and $\#\mathcal{C} \times 1$ subvector of $F_S(\int m_S(x,\theta)dQ_{\alpha_\beta}(x))$ that correspond to the constraints indexed by $\mathcal{C}$. Since $\lambda^{(i)}(p_{\alpha_\beta}, Q_{\alpha_\beta}) = 0$ for all $i \notin \mathcal{C}$ by (C.4), we then have:

$$F_A^{\mathcal{C}}(\int m_A(x)dP(x))'\lambda^{\mathcal{C}}(p^*, P) = -p^* \ , \tag{C.7}$$

by results (C.4), (C.6) and Lemmas A.5 and A.12. Moreover, note that by definition of $\mathcal{C}$ we also obtain that:

$$F_A^{\mathcal{C}}(\int m_A(x)dQ_{\alpha_\beta}(x))\theta_{\alpha_\beta} = -F_S^{\mathcal{C}}(\int m_S(x,\theta_{\alpha_\beta})dQ_{\alpha_\beta}(x)) \ . \tag{C.8}$$

Moreover, since $\mathcal{S}_i = \emptyset$ for all $i \in \mathcal{C}$ by (C.5), (C.8) is a linear equation in θ_{α_β}, and by $Q_{\alpha_\beta} \notin \mathbf{M_L}$ satisfying Assumption 3.6(iv) we must have $\#\mathcal{C} < d_\theta$, for otherwise (C.8) would have a unique solution in θ and (C.3) would imply $\Xi(p_{\alpha_\beta}, Q_{\alpha_\beta})$ is a singleton. Thus, while (C.4), (C.8) and Lemma A.5 imply $\mathcal{C} \subseteq \mathcal{A}(\theta^*, P)$, we may also conclude from $\#\mathcal{C} < d_\theta$ and $\Xi(p^*, P)$ being a singleton by $(p^*, P) \in \mathbb{C} \times \mathbf{M_L}$, that we also have:

$$\mathcal{A}(\theta^*, P) \setminus \mathcal{C} \neq \emptyset \ . \tag{C.9}$$

In what follows we aim to establish a contradiction by showing that P will not satisfy Assumption 3.6(iv) at the point $\theta^* \in \Theta_0(P)$. To this end, for notational convenience we first define the sets:

$$K_i \equiv \{\theta \in \Theta : F^{(i)}(\int m(x,\theta)dP(x)) \leq 0\} \qquad E_i \equiv \{\theta \in \Theta : F^{(i)}(\int m(x,\theta)dP(x)) = 0\} \ . \tag{C.10}$$

Next, note that $\Xi(p^*, P) = \{\theta^*\}$ and convexity of $F^{(i)}(\int m(x,\cdot)dP(x)) : \Theta \to \mathbf{R}$ for all $1 \leq i \leq d_F$ imply:

$$\{\theta^*\} = \{\bigcap_{1\leq i\leq d_F} K_i\} \cap \{\theta \in \Theta : \langle p^*, \theta\rangle = \nu(p^*, \Theta_0(P))\} = \{\bigcap_{i\in\mathcal{A}(\theta^*,P)} K_i\} \cap \{\theta \in \Theta : \langle p^*, \theta\rangle = \nu(p^*, \Theta_0(P))\} \ . \tag{C.11}$$

Moreover, also note $\mathcal{C} \subseteq \mathcal{A}(\theta^*, P)$ implies $F_A^{\mathcal{C}}(\int m_A(x)dP(x))\theta^* = -F_S^{\mathcal{C}}(\int m_S(x,\theta^*)dP(x))$, and hence by (C.7):

$$\lambda^{\mathcal{C}}(p^*, P)'F_S^{\mathcal{C}}(\int m_S(x,\theta^*)dP(x)) = \langle p^*, \theta^*\rangle = \nu(p^*, \Theta_0(P)) \ . \tag{C.12}$$

Since $\mathcal{S}_i = \emptyset$ for all $i \in \mathcal{C}$, results (C.7) and (C.12) imply $\{\bigcap_{i\in\mathcal{C}} E_i\} \subseteq \{\theta \in \Theta : \langle p^*, \theta\rangle = \nu(p^*, \Theta_0(P))\}$, which yields

$$\{\theta^*\} = \{\bigcap_{i\in\mathcal{A}(\theta^*,P)\setminus\mathcal{C}} K_i\} \cap \{\bigcap_{i\in\mathcal{C}} E_i\} \ , \tag{C.13}$$

due to (C.9), (C.11), and $E_i \subseteq K_i$. Next, let $\iota : \mathcal{A}(\theta^*, P) \setminus \mathcal{C} \to \{1, \ldots, \#\mathcal{A}(\theta^*, P) \setminus \mathcal{C}\}$ be a bijection, and define:

$$j^* \equiv \min_{1 \leq j \leq \#\mathcal{A}(\theta^*,P)\setminus\mathcal{C}} j : \{\bigcap_{i \in \mathcal{A}(\theta^*,P)\setminus\mathcal{C}:\iota(i)\leq j} K_i\} \cap \{\bigcap_{i\in\mathcal{C}} E_i\} \text{ is a singleton} \ , \tag{C.14}$$

where we note j^* is well defined due to (C.13), and $\{\bigcap_{i\in\mathcal{C}} E_i\}$ not being singleton due to $\#\mathcal{C} < d_\theta$ and $F^{(i)}(\int m(x, \cdot)dP(x)) : \Theta \to \mathbf{R}$ being linear for all $i \in \mathcal{C}$. Thus, from (C.10), (C.14) and setting $i^* \equiv \iota^{-1}(j^*) \in \mathcal{A}(\theta^*, P)$ we conclude:[14]

$$\{\theta^*\} = \arg\min_{\theta\in\Theta} \Big\{F^{(i^*)}(\int m(x,\theta)dP(x)) \text{ s.t. } \theta \in \{\bigcap_{i:\iota(i)\leq j^*-1} K_i\} \cap \{\bigcap_{i\in\mathcal{C}} E_i\}\Big\} \ . \tag{C.15}$$

However, since the constraint set is not a singleton, it follows that for each i such that $\iota(i) \leq j^* - 1$, either $F^{(i)}(\int m(x, \theta)dP(x))$ is linear in θ (if $i \in \mathcal{I}_L$), or $F^{(i)}(\int m(x, \theta_i)dP(x)) < 0$ for some $\theta_i \in \{\bigcap_{i:\iota(i)\leq j^*-1} K_i\} \cap \{\bigcap_{i\in\mathcal{C}} E_i\}$ (if $i \in \mathcal{I}_S$). It follows that (C.15) is an ordinary convex problem satisfying a primal qualification constraint, and by Theorem 28.2 in Rockafellar (1970) that there exist Kuhn-Tucker vectors such that:

$$\{\theta^*\} = \arg\min_{\theta\in\Theta} \Big\{F^{(i^*)}(\int m(x,\theta)dP(x)) + \sum_{i:\iota(i)\leq j^*-1} \gamma_i F^{(i)}(\int m(x,\theta)dP(x)) + \sum_{i\in\mathcal{C}} \pi_i F^{(i)}(\int m(x,\theta)dP(x))\Big\} \ . \tag{C.16}$$

Finally, we observe that since $\theta^* \in \Theta_0(P) \subseteq \Theta^o$ by Assumption 3.6(i), result (C.16) and Corollary A.1 imply:

$$-\nabla_\theta F^{(i^*)}(\int m(x,\theta^*)dP(x)) = \sum_{i:\iota(i)\leq j^*-1} \gamma_i \nabla_\theta F^{(i)}(\int m(x,\theta^*)dP(x)) + \sum_{i\in\mathcal{C}} \pi_i \nabla_\theta F^{(i)}(\int m(x,\theta^*)dP(x)) \ . \tag{C.17}$$

Thus, we reach the desired contradiction that $P \in \mathbf{M_L}$ violates Assumption 3.6(iv). ■

Lemma C.3. *If Assumptions 3.2, 3.4, 3.5, 4.2 hold, $P \in \mathbf{P_L}$, $\mathbf{S_L} \equiv \{h \in L^2_\mu : h = \sqrt{dQ/d\mu}$ for some $Q \in \mathbf{P_L}\}$, and $\mathbb{C}$ is compact, then the tangent space of $\mathbf{S_L}$ at $s = \sqrt{dP/d\mu}$ is $\dot{\mathbf{S}}_\mathbf{L} = \{h \in L^2_\mu : \int h(x)s(x)d\mu(x) = 0\}$.*

Proof: The claim follows immediately from Lemmas A.16 and C.2. ■

Lemma C.4. *If Assumptions 3.2, 3.4, 3.5, 4.2(i) hold and $\mathbb{C}$ is compact, then the mapping $\rho_L : \mathbf{P_L} \to \mathcal{C}(\mathbb{C})$ pointwise defined by $\rho_L(P) = \nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$ is weak-pathwise differentiable at any $P \in \mathbf{P_L}$. Moreover, for $s \equiv \sqrt{dP/d\mu}$, $\lambda(p, Q)$ (as in* (A.29)*), and $\{\theta^*(p)\} = \Xi(p, P)$ (as in* (A.24)*), the derivative $\dot\rho_L : \dot{\mathbf{S}}_\mathbf{L} \to \mathcal{C}(\mathbb{C})$ satisfies:*

$$\dot\rho_L(\dot h_0)(p) = 2\lambda(p, P)'\nabla F(\int m(x, \theta^*(p))dP(x)) \int m(x, \theta^*(p))\dot h_0(x)s(x)d\mu(x) \ .$$

Proof: First note $\dot\rho_L(\dot h_0) \in \mathcal{C}(\mathbb{C})$ for any $\dot h_0 \in \dot{\mathbf{S}}_\mathbf{L}$ by Lemma C.1. In addition, $\dot\rho_L : \dot{\mathbf{S}}_\mathbf{L} \to \mathcal{C}(\mathbb{C})$ is linear, and bounded since by Lemma A.11, $P \in \mathbf{P_L}$ satisfying Assumption 3.6(iii), and Assumptions 3.4(i) and 3.5(ii) we have:

$$\begin{aligned}\sup_{\|\dot h_0\|_{L^2_\mu}=1} \sup_{p\in\mathbb{C}} |\dot\rho_L(\dot h_0)(p)| \\ \leq \sup_{\|\dot h_0\|_{L^2_\mu}} \sup_{p\in\mathbb{C}} \{2\|\lambda(p,P)\| \times \sup_{v\in V_0} \|\nabla F(v)\|_F \times \sqrt{d_m} \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \|m(x,\theta)\| \times \|\dot h_0\|_{L^2_\mu} \times \|s\|_{L^2_\mu}\} < \infty \ . \end{aligned} \tag{C.18}$$

Finally, note that for any curve $\eta \mapsto P_\eta$ in $\mathbf{P_L}$ with $h_0 = s$ and all finite Borel measures B on $\mathbb{C}$, the mean value theorem, the dominated convergence theorem and Lemma C.1 allow us to conclude that:

$$\lim_{\eta_0\to 0} \int_{\mathbb{C}} \{\frac{\nu(p, \Theta_0(P_{\eta_0})) - \nu(p, \Theta_0(P))}{\eta_0} - \dot\rho_L(\dot h_0)(p)\}dB(p) = 0 \ , \tag{C.19}$$

(see (A.57)). Since (C.19) verifies $\dot\rho_L : \dot{\mathbf{S}}_\mathbf{L} \to \mathcal{C}(\mathbb{C})$ is the weak derivative of $\rho_L : \mathbf{P_L} \to \mathcal{C}(\mathbb{C})$, the Lemma follows. ■

Theorem C.1. *Let Assumptions 3.1, 3.2, 3.4, 3.5, 4.2 hold, $P \in \mathbf{P_L}$ and $\mathbb{C}$ be compact. For each $\theta_1, \theta_2 \in \Theta$, let $H(\theta_1)$ and $\Omega(\theta_1, \theta_2)$ be as in Theorem 3.2, $\{\theta^*(p)\} = \Xi(p, P)$ (as in* (A.24)*) and define $\rho_L : \mathbf{P_L} \to \mathcal{C}(\mathbb{C})$ by*

[14] Here $\{\bigcap_{i\in\emptyset} K_i\} \cap \{\bigcap_{i\in\mathcal{C}} E_i\}$ should be understood to equal $\{\bigcap_{i\in\mathcal{C}} E_i\}$.

$\rho_L(P) \equiv \nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$. *The inverse information covariance functional for estimating* $\rho_L(P)$ *is then given by:*

$$I^{-1}(p_1, p_2) = \lambda(p_1, P)' H(\theta^*(p_1)) \Omega(\theta^*(p_1), \theta^*(p_2)) H(\theta^*(p_2))' \lambda(p_2, P) \ . \tag{C.20}$$

Proof: As in the proof of Theorem 3.2, we closely follow Chapter 5.2 in Bickel et al. (1993). Let $\mathbf{B} \equiv \mathcal{C}(\mathbb{C})$ and $\mathbf{B}^*$ denote the set of finite Borel measures on $\mathbb{C}$, which by Corollary 14.15 in Aliprantis and Border (2006) is the dual space of $\mathbf{B}$. For $s \equiv \sqrt{dP/d\mu}$ then define $\dot{\rho}_L^T : \mathbf{B}^* \to \dot{\mathbf{S}}_\mathbf{L}$ pointwise by:

$$\dot{\rho}_L^T(B)(x) \equiv \int_{\mathbb{C}} 2\lambda(p, P)' H(\theta^*(p)) \{m(x, \theta^*(p)) - E[m(X_i, \theta^*(p))]\} s(x) dB(p) \ , \tag{C.21}$$

noting the integrand is indeed measurable by arguing as in (A.59) and exploiting $p \mapsto \theta^*(p)$ is continuous on $\mathbb{C}$ due to Lemma A.7 and $\Xi(p, P)$ being a singleton for all $p \in \mathbb{C}$ due to $P \in \mathbf{P_L}$. For any $B \in \mathbf{B}^*$ let $\Gamma(B)$ denote the finite Borel measure on $\mathbb{S}^{d_\theta}$ given by $\Gamma(B)(A) = B(A \cap \mathbb{C})$ for any Borel set $A \subseteq \mathbb{S}^{d_\theta}$. Noting that $\dot{\rho}_L^T(B) = \dot{\rho}^T(\Gamma(B))$, it then follows from Lemma C.3 and results (A.60)-(A.62) that $\dot{\rho}_L^T : \mathbf{B}^* \to \dot{\mathbf{S}}_\mathbf{L}$ is the adjoint of $\dot{\rho}_L : \dot{\mathbf{S}}_\mathbf{L} \to \mathbf{B}$. Lemmas C.3 and C.4 and Theorem 5.2.1 in Bickel et al. (1993) then establish the Theorem. ■

Lemma C.5. *Let Assumptions 3.2, 3.4, 3.5, 4.2 hold,* $\mathbb{C}$ *be compact,* $P \in \mathbf{P_L}$ *and* $Q_\tau \equiv \tau Q + (1 - \tau)P$ *for any* $Q \in \mathbf{M}$. *Then, there is a* $N(P) \subseteq \mathbf{M}$ *such that for all* $(Q, p, \tau_0) \in N(P) \times \mathbb{C} \times (0, 1)$*:*

$$\frac{\partial}{\partial \tau} \nu(p, \Theta_0(Q_\tau))\Big|_{\tau=\tau_0} = \lambda(p, Q_{\tau_0})' \nabla F(\int m(x, \theta^*) dQ_{\tau_0}(x)) \int m(x, \theta^*)(dQ(x) - dP(x)) \quad \text{where } \{\theta^*\} = \Xi(p, Q_{\tau_0}) \ .$$

Proof: By Lemmas B.2 and C.2 there is a $N(P) \subseteq \mathbf{M}$ that is convex and contained in $\mathbf{M_L}$ (as in (C.2)). Hence, if $Q \in N(P) \subseteq \mathbf{M_L}$, then $Q_\tau \in \mathbf{M_L}$ for all $\tau \in (0, 1)$ which together with Assumption 3.5, Lemma A.9 and Corollary 5 in Milgrom and Segal (2002) imply that for any $(Q, p) \in N(P) \times \mathbb{C}$ the function $\tau \mapsto \nu(p, \Theta_0(Q_\tau))$ is directionally differentiable with right and left derivatives given by:

$$\frac{\partial}{\partial \tau_+} \nu(p, \Theta_0(Q_\tau))\Big|_{\tau=\tau_0} = \max_{\theta^* \in \Xi(p, Q_{\tau_0})} \lambda(p, Q_{\tau_0})' \nabla F(\int m(x, \theta^*) dQ_{\tau_0}(x)) \int m(x, \theta^*)(dQ(x) - dP(x)) \tag{C.22}$$

$$\frac{\partial}{\partial \tau_-} \nu(p, \Theta_0(Q_\tau))\Big|_{\tau=\tau_0} = \min_{\theta^* \in \Xi(p, Q_{\tau_0})} \lambda(p, Q_{\tau_0})' \nabla F(\int m(x, \theta^*) dQ_{\tau_0}(x)) \int m(x, \theta^*)(dQ(x) - dP(x)) \tag{C.23}$$

(see also (B.4)-(B.5)). However, since $Q_{\tau_0} \in N(P) \subseteq \mathbf{M_L}$ for all $\tau_0 \in (0, 1)$, it follows that for any $p \in \mathbb{C}$ the correspondence $\Xi(p, Q_{\tau_0})$ is singleton valued. We conclude (C.22) and (C.23) agree, and the Lemma follows. ■

Lemma C.6. *Let Assumptions 3.1, 3.2, 3.4, 3.5, 4.2 hold,* $\mathbb{C}$ *be compact,* $P \in \mathbf{P_L}$ *and* $\{\theta^*(p)\} = \Xi(p, P)$. *Then:*

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

Proof: By Lemma B.2 we may restrict attention to convex neighborhoods, so that if $\hat{P}_n \in N(P)$ then $\hat{P}_{n,\tau} \equiv \tau \hat{P}_n + (1 - \tau)P \in N(P)$ for all $\tau \in [0, 1]$. Hence, Lemmas A.7 and B.5 imply $\Xi(p, \hat{P}_{n,\tau})$ is well defined for all $\tau \in [0, 1]$ with probability tending to one. Moreover, since $P \in \mathbf{P_L}$ implies $\Xi(p, P)$ is singleton valued for all $p \in \mathbb{C}$, we obtain:

$$\liminf_{n \to \infty} P(\sup_{p \in \mathbb{C}} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \|\theta - \theta^*(p)\| > \epsilon) = 0 \tag{C.24}$$

for any $\epsilon > 0$ due to Lemmas A.7, B.4 and B.5. Thus, since $p \mapsto \lambda(p, P)$ and $p \mapsto H(\theta^*(p))$ are uniformly bounded on $\mathbb{C}$ by Lemma A.11, Assumption 3.5 and $P \in \mathbf{P_L}$ satisfying Assumption 3.6(iii), we obtain

$$\sup_{p \in \mathbb{C}} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \|\sqrt{n} \lambda(p, P)' H(\theta^*(p)) \int (m(x, \theta) - m(x, \theta^*(p)))(d\hat{P}_n(x) - dP(x))\| = o_p(1) \tag{C.25}$$

due to result (C.24) and Lemma B.1 (see also (B.27)-(B.28)). Additionally, since Θ is compact, result (A.3) implies $\theta \mapsto \int m(x, \theta) dP(x)$ is uniformly continuous on Θ, and we therefore obtain from Lemma B.1 that (see also (B.23)):

$$\sup_{p \in \mathbb{C}} \sup_{\tau \in [0,1]} \sup_{\theta \in \Xi(p, \hat{P}_{n,\tau})} \|\int m(x, \theta) d\hat{P}_{n,\tau}(x) - \int m(x, \theta^*(p)) dP(x)\| = o_p(1) \ . \tag{C.26}$$

Further note $\nabla F(\int m(x,\theta)dP(x))$ is uniformly bounded in $\theta \in \Theta$ by Assumption 3.5 and $P \in \mathbf{P_L}$ satisfying Assumption 3.6(iii), while $\lambda(p,P)$ is uniformly bounded on $\mathbb{C}$ by Lemma A.11. Therefore, $v \mapsto \nabla F(v)$ being uniformly continuous on V_0 by Assumption 3.5(ii), together with Lemmas A.2 and B.5 and results (B.31) and (C.26) yield:

$$\sup_{p\in\mathbb{C}} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(p,\hat{P}_{n,\tau})} \|\lambda(p,\hat{P}_{n,\tau})'\nabla F(\int m(x,\theta)d\hat{P}_{n,\tau}(x)) - \lambda(p,P)'\nabla F(\int m(x,\theta^*(p))dP(x))\| = o_p(1) \ . \quad \text{(C.27)}$$

Finally, employing the mean value theorem, which is valid by Lemmas B.2, B.5 and C.5, we obtain uniformly in $p \in \mathbb{C}$ that for some $\tau_0 : \mathbb{C} \to (0,1)$ and $\tilde{\theta} : \mathbb{C} \to \Theta$ with $\tilde{\theta}(p) \in \Xi(p,\hat{P}_{n,\tau_0(p)})$ for all $p \in \mathbb{C}$:

$$\begin{aligned}\sqrt{n}\{\nu(p,\Theta_0(\hat{P}_n)) - \nu(p,\Theta_0(P))\} \\ &= \sqrt{n}\lambda(p,\hat{P}_{n,\tau_0(p)})'\nabla F(\int m(x,\tilde{\theta}(p))d\hat{P}_{n,\tau_0(p)}(x)) \int m(x,\tilde{\theta}(p))(d\hat{P}_n(x) - dP(x)) + o_p(1) \\ &= \sqrt{n}\lambda(p,P)'H(\theta^*(p)) \int m(x,\theta^*(p))(d\hat{P}_n(x) - dP(x)) + o_p(1) \ , \qquad \text{(C.28)}\end{aligned}$$

where the second equality follows from results (B.25), (C.25) and (C.27). ■

Proof of Theorem 4.3: We first show the class $\mathcal{F} \equiv \{f : \mathcal{X} \to \mathbf{R} : f(x) = \lambda(p,P)'H(\theta^*(p))m(x,\theta^*(p))$ for some $p \in \mathbb{C}\}$ is Donsker in $\mathcal{C}(\mathbb{C})$. To this end note that $p \mapsto \lambda(p,P)'H(\theta^*(p))$ and $p \mapsto \theta^*(p)$ are continuous in $p \in \mathbb{C}$ due to Lemmas A.7 and A.12, result (A.3), Assumption 3.5 and $P \in \mathbf{P_L}$ satisfying Assumption 3.6(iii). Thus, it follows from Assumption 3.4(i)-(ii) that $f \in \mathcal{F}$ are uniformly bounded, and that the empirical process belongs to $\mathcal{C}(\mathbb{C})$. Convergence of the marginals is then immediate, while for any sequence $\delta_n \downarrow 0$ we obtain

$$\sup_{p_1,p_2\in\mathbb{C}:\|p_1-p_2\|\le\delta_n} |\sqrt{n}\int(m(x,\theta^*(p_1)) - m(x,\theta^*(p_2)))(d\hat{P}_n(x) - dP(x))| = o_p(1) \ , \quad \text{(C.29)}$$

due to Lemma B.1 and continuity of $p \mapsto \theta^*(p)$ on $\mathbb{C}$. The class $\mathcal{F}$ being Donsker then follows from (C.29), Lemma B.1 and $p \mapsto \lambda(p,P)'H(\theta^*(p))$ being uniformly continuous and bounded on $\mathbb{C}$ by compactness. Theorem 18.1 in Kosorok (2008) and Lemma C.6 then imply $\{\nu_{|\mathbb{C}}(\cdot,\Theta_0(\hat{P}_n))\}$ is a regular estimator of $\nu_{|\mathbb{C}}(\cdot,\Theta_0(P))$. The Theorem then follows from the influence function of $\{\nu_{|\mathbb{C}}(\cdot,\Theta_0(\hat{P}_n))\}$ being efficient by Lemma C.6 and Theorem C.1. ■

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

The proofs of all Theorems in this section are self contained, and do not require auxiliary Lemmas or results.

Proof of Theorem 5.1: 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:

$$\sup_{f\in BL_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} |E[f(G_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| = o_p(1) \ , \quad \text{(D.1)}$$

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

$$\begin{aligned}&\sup_{p\in\mathbb{S}^{d_\theta}} \|\sqrt{n}\int w\{m(x,\hat{\theta}(p)) - \int m(x,\hat{\theta}(p))d\hat{P}_n(x)\}d\hat{P}_n^W(x,w)\| \\ &\le \sup_{\theta\in\Theta} \|\sqrt{n}\int wm(x,\theta)d\hat{P}_n^W(x,w)\| + \sup_{(x,\theta)\in(\mathcal{X}\times\Theta)} \|m(x,\theta)\| \times |\sqrt{n}\int wd\hat{P}_n^W(x,w)| = O_p(1) \quad \text{(D.2)}\end{aligned}$$

due to $W_i \perp X_i$, $E[W_i] = 0$ by Assumption 5.1(ii) and $(x,\theta) \mapsto m(x,\theta)$ being uniformly bounded by Assumption 3.4(i). Next, let $\Pi_p : \Theta \to \mathbf{R}^{d_\theta}$ be as in Lemma B.7, and note Lemmas B.5 and B.7 imply uniformly in $p \in \mathbb{S}^{d_\theta}$

$$\begin{aligned}\lambda(p,P)'\nabla F_S(\int m_S(x,\hat{\theta}(p))d\hat{P}_n(x)) \int m_S(x,\hat{\theta}(p))d\hat{P}_n(x) \\ = \lambda(p,P)'\nabla F_S(\int m_S(x,\Pi_p\hat{\theta}(p)))d\hat{P}_n(x)) \int m_S(x,\Pi_p\hat{\theta}(p))d\hat{P}_n(x) + o_p(1) \\ = \lambda(p,P)'\nabla F_S(\int m_S(x,\theta^*(p)))dP(x)) \int m_S(x,\Pi_p\hat{\theta}(p))d\hat{P}_n(x) + o_p(1) \\ = \lambda(p,P)'\nabla F_S(\int m_S(x,\theta^*(p)))dP(x)) \int m_S(x,\theta^*(p))dP(x) + o_p(1) ,\end{aligned} \quad \text{(D.3)}$$

where the second equality follows from (B.20), Assumption 3.4(i) and (B.24), while the third equality results from Lemma A.11, Assumption 3.5(ii), $P \in \mathbf{P}$ satisfying Assumption 3.6(iii) and result (B.23). Therefore, results (B.31), Assumption 3.5(ii), Lemmas A.2 and B.5, and result (D.2) yield uniformly in $p \in \mathbb{S}^{d_\theta}$

$$\begin{aligned}&\sqrt{n}\lambda(p,\hat{P}_n)'\nabla F(\int m(x,\hat{\theta}(p))d\hat{P}_n(x)) \int w\{m(x,\hat{\theta}(p)) - \int m(x,\hat{\theta}(p))d\hat{P}_n(x)\}d\hat{P}_n^W(x,w) \\ &= \sqrt{n}\lambda(p,P)'\nabla F(\int m(x,\hat{\theta}(p))d\hat{P}_n(x)) \int w\{m(x,\hat{\theta}(p)) - \int m(x,\hat{\theta}(p))d\hat{P}_n(x)\}d\hat{P}_n^W(x,w) + o_p(1) \\ &= \sqrt{n}\lambda(p,P)'\nabla F(\int m(x,\theta^*(p))dP(x)) \int w\{m(x,\theta^*(p)) - \int m(x,\theta^*(p))dP(x)\}d\hat{P}_n^W(x,w) + o_p(1) ,\end{aligned} \quad \text{(D.4)}$$

where the second equality follows from $A\theta - \int A\theta d\hat{P}_n(x) = 0$, $E[W_i] = 0$ and $W_i \perp X_i$ by Assumption 5.1, Lemma B.8 and result (D.3). Next, define the process G_n^* to be pointwise given by:

$$\bar{G}_n^*(p) \equiv \sqrt{n}\lambda(p,P)'H(\theta^*(p)) \int w\{m(x,\theta^*(p)) - \int m(x,\theta^*(p))dP(x)\}d\hat{P}_n^W(x,w) , \quad \text{(D.5)}$$

and note arguments identical to those in (A.50)-(A.51) imply that $\bar{G}_n^* \in \mathcal{C}(\mathbb{S}^{d_\theta})$ 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:

$$\sup_{f \in BL_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} |E[f(\bar{G}_n^*) - f(G_n^*)|\{X_i\}_{i=1}^n]| \leq \eta P(\|\bar{G}_n^* - G_n^*\|_\infty \leq \eta|\{X_i\}_{i=1}^n) + 2P(\|\bar{G}_n^* - G_n^*\|_\infty > \eta|\{X_i\}_{i=1}^n) . \quad \text{(D.6)}$$

However, from (D.4), it follows that $P(\|\bar{G}_n^* - G_n^*\|_\infty > \eta|\{X_i\}_{i=1}^n) = o_p(1)$, and hence since η in (D.6) is arbitrary:

$$\sup_{f \in BL_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} |E[f(G_n^*)|\{X_i\}_{i=1}^n] - E[f(\bar{G}_n^*)|\{X_i\}_{i=1}^n]| = o_p(1) . \quad \text{(D.7)}$$

To conclude, we note that by Lemma B.10 and Theorem 2.9.6 in van der Vaart and Wellner (1996), we have:

$$\sup_{f \in BL_1(\mathcal{C}(\mathbb{S}^{d_\theta}))} |E[f(\bar{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| = o_p(1) , \quad \text{(D.8)}$$

and therefore results (D.7) and (D.8) verify (D.1) which establishes the claim of the Theorem. ■

Proof of Theorem 5.2: Let $\bar{G}_n^*$ be defined as in (D.5) and note that by (D.4) $\|\bar{G}_n^* - G_n^*\|_\infty = o_p(1)$ unconditionally. Define a mapping $\Gamma : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathcal{C}(\mathbb{S}^{d_\theta})$ pointwise by $\Gamma(f) = \Upsilon \circ f$. The continuous mapping theorem then yields:

$$|\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) - \sup_{p \in \hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p))| \leq \sup_{p \in \mathbb{S}^{d_\theta}} |\Upsilon(G_n^*(p)) - \Upsilon(\bar{G}_n^*(p))| = \|\Gamma(G_n^*) - \Gamma(\bar{G}_n^*)\|_\infty = o_p(1) . \quad \text{(D.9)}$$

Next, let $\hat{p}^* \in \arg\max_{p \in \hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p))$ which is well defined by Assumption 5.2(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:

$$\sup_{p \in \hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p)) - \sup_{p \in \Psi_0} \Upsilon(\bar{G}_n^*(p)) \leq \Upsilon(\bar{G}_n^*(\hat{p}^*)) - \Upsilon(\bar{G}_n^*(\Pi_{\Psi_0}\hat{p}^*)) \leq \sup_{\|p - \tilde{p}\| \leq d_H(\hat{\Psi}_n, \Psi_0)} |\Upsilon(\bar{G}_n^*(p)) - \Upsilon(\bar{G}_n^*(\tilde{p}))| . \quad \text{(D.10)}$$

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 conclude:

$$|\sup_{p \in \hat{\Psi}_n} \Upsilon(\bar{G}_n^*(p)) - \sup_{p \in \Psi_0} \Upsilon(\bar{G}_n^*(p))| \leq \sup_{\|p - \tilde{p}\| \leq d_H(\hat{\Psi}_n, \Psi_0)} |\Upsilon(\bar{G}_n^*(p)) - \Upsilon(\bar{G}_n^*(\tilde{p}))| . \quad \text{(D.11)}$$

By Assumption 5.1, Lemma B.10 and Theorem 2.9.2 in van der Vaart and Wellner (1996), $\bar{G}_n^* \xrightarrow{L} \bar{\mathbb{G}}$ (unconditionally) for some tight Gaussian process $\bar{\mathbb{G}}$ 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 $\mathbf{R}$. Next, fix $\eta > 0$, $\epsilon > 0$ and note there then is a constant $K > 0$ such that:

$$\limsup_{n\to\infty} P(\sup_{p\in\mathbb{S}^{d_\theta}} |\bar{G}_n^*(p)| > K) < \eta \ . \tag{D.12}$$

By Assumption 5.2(i), $\Upsilon : \mathbf{R} \to \mathbf{R}$ 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:

$$\limsup_{n\to\infty} P(\sup_{\|p-\tilde{p}\|\le d_H(\hat{\Psi}_n,\Psi_0)} |\Upsilon(\bar{G}_n^*(p)) - \Upsilon(\bar{G}_n^*(\tilde{p}))| > \epsilon)$$
$$\le \limsup_{n\to\infty} P(\sup_{\|p-\tilde{p}\|\le d_H(\hat{\Psi}_n,\Psi_0)} |\bar{G}_n^*(p) - \bar{G}_n^*(\tilde{p})| > \delta_0) + \limsup_{n\to\infty} P(\sup_{p\in\mathbb{S}^{d_\theta}} |\bar{G}_n^*(p)| > K) \ . \tag{D.13}$$

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:

$$\limsup_{n\to\infty} P(\sup_{\|p-\tilde{p}\|\le d_H(\hat{\Psi}_n,\Psi_0)} |\bar{G}_n^*(p) - \bar{G}_n^*(\tilde{p})| > \delta_0)$$
$$\le \limsup_{n\to\infty} P(\sup_{\|p-\tilde{p}\|\le \gamma_0} |\bar{G}_n^*(p) - \bar{G}_n^*(\tilde{p})| > \delta_0) + \limsup_{n\to\infty} P(d_H(\hat{\Psi}_n, \Psi_0) > \gamma_0) < \eta \ , \tag{D.14}$$

due to $d_H(\hat{\Psi}_n, \Psi_0) = o_p(1)$ by hypothesis. Since ϵ, η were arbitrary, combining (D.9)-(D.14) we then obtain:

$$\sup_{p\in\hat{\Psi}_n} \Upsilon(G_n^*(p)) = \sup_{p\in\Psi_0} \Upsilon(\bar{G}_n^*(p)) + o_p(1) \ . \tag{D.15}$$

Therefore, for $BL_1(\mathbf{R})$ as in (D.1), arguing as in (D.7) and using Theorem 5.1 and Theorem 10.8 in Kosorok (2008):

$$\sup_{f\in BL_1(\mathbf{R})} |E[f(\sup_{p\in\hat{\Psi}_n} \Upsilon(G_n^*(p)))|\{X_i\}_{i=1}^n] - E[f(\sup_{p\in\Psi_0} \Upsilon(\mathbb{G}_0(p)))]|$$
$$\le \sup_{f\in BL_1(\mathbf{R})} |E[f(\sup_{p\in\Psi_0} \Upsilon(\bar{G}_n^*(p)))|\{X_i\}_{i=1}^n] - E[f(\sup_{p\in\Psi_0} \Upsilon(\mathbb{G}_0(p)))]| + o_p(1) = o_p(1) \ . \tag{D.16}$$

To conclude, observe that result (D.16) together with Lemma 10.11 in Kosorok (2008) imply that:

$$P(\sup_{p\in\hat{\Psi}_n} \Upsilon(G_n^*(p)) \le t|\{X_i\}_{i=1}^n) = P(\sup_{p\in\Psi_0} \Upsilon(\mathbb{G}_0(p)) \le t) + o_p(1) \tag{D.17}$$

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 $\tilde{\epsilon} \le \epsilon$ such that $c_{1-\alpha} \pm \tilde{\epsilon}$ are also continuity points and in addition:

$$P(\sup_{p\in\Psi_0} \Upsilon(\mathbb{G}_0(p)) \le c_{1-\alpha} - \tilde{\epsilon}) < 1 - \alpha < P(\sup_{p\in\Psi_0} \Upsilon(\mathbb{G}_0(p)) \le c_{1-\alpha} + \tilde{\epsilon}) \ , \tag{D.18}$$

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:

$$A_n \equiv \{P(\sup_{p\in\hat{\Psi}_n} \Upsilon(G_n^*(p)) \le c_{1-\alpha} - \tilde{\epsilon}|\{X_i\}_{i=1}^n) < 1 - \alpha < P(\sup_{p\in\hat{\Psi}_n} \Upsilon(G_n^*(p)) \le c_{1-\alpha} + \tilde{\epsilon}|\{X_i\}_{i=1}^n)\} \tag{D.19}$$

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 (D.17) yields that:

$$\liminf_{n\to\infty} P(|\hat{c}_{1-\alpha} - c_{1-\alpha}| \le \epsilon) \ge \liminf_{n\to\infty} P(A_n) = 1 \ , \tag{D.20}$$

which establishes the claim of the Theorem. ■

Proof of Theorem 5.3: Since support functions are continuous, 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 4.1 and Corollary 1.10 in Li et al. (2002) imply that:

$$d_H(\Theta_0(P), \hat{\Theta}_n) = O_p(n^{-\frac{1}{2}}) \ . \tag{D.21}$$

In turn, result (D.21) and Lemma B.10 in Kaido (2010) yield $d_H(\hat{\mathfrak{M}}_n(\theta), \mathfrak{M}(\theta)) = o_p(1)$. Therefore, Assumption 5.2 is satisfied with $\mathfrak{M}(\theta) = \Psi_0$ and $\hat{\mathfrak{M}}_n(\theta) = \hat{\Psi}_n$. Moreover, 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 and $\text{Var}\{\mathbb{G}_0(p)\} > 0$ for all $p \in \mathfrak{M}(\theta)$ by hypothesis, we obtain for any $p_0 \in \mathfrak{M}(\theta)$ that:

$$P(\sup_{p\in\mathfrak{M}(\theta)} |-\mathbb{G}_0(p)|_+ \le 0) \le P(-\mathbb{G}_0(p_0) \le 0) = 0.5 \ . \tag{D.22}$$

Therefore, $\alpha < 0.5$ 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 5.2 it then follows that $\hat{c}_{1-\alpha}(\theta) = c_{1-\alpha}(\theta) + o_p(1)$.

Suppose $\theta \in \Theta_0(P)^o$. Then result (D.21) implies that with probability tending to one $\theta \in \hat{\Theta}_n^o$. Therefore, $J_n(\theta) = 0$ with probability tending to one, and since $\hat{c}_{1-\alpha}(\theta) \xrightarrow{p} c_{1-\alpha}(\theta) > 0$, we conclude:

$$\liminf_{n\to\infty} P(J_n(\theta) \le \hat{c}_{1-\alpha}(\theta)) = 1 \ . \tag{D.23}$$

Suppose on the other hand that $\theta \in \partial\Theta_0(P)$. Theorem 4.1 and Lemma B.9 in Kaido (2010) then imply that:

$$J_n(\theta) \xrightarrow{L} \sup_{p\in\mathfrak{M}(\theta)} |-\mathbb{G}_0(p)|_+ \ . \tag{D.24}$$

Therefore, since $\hat{c}_{1-\alpha}(\theta) \xrightarrow{p} c_{1-\alpha}(\theta)$ and the cdf of $\sup_{p\in\mathfrak{M}(\theta)} |-\mathbb{G}_0(\theta)|_+$ is continuous at $c_{1-\alpha}(\theta)$, (D.24) yields:

$$\lim_{n\to\infty} P(J_n(\theta) \le \hat{c}_{1-\alpha}(\theta)) = P(\sup_{p\in\mathfrak{M}(\theta)} |-\mathbb{G}_0(p)|_+ \le c_{1-\alpha}(\theta)) = 1-\alpha \ , \tag{D.25}$$

which establishes the claim of the Theorem. ■

Proof of Theorem 5.4: We first study the behavior of $\{\pi_n^*\}$. To this end define the functional $\psi : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}$ to be pointwise given by $\psi(f) = \sup_{p\in\mathbb{S}^{d_\theta}}\{\nu(p, \{\theta_0\}) - f(p)\}$, and the event $A_n \equiv \{\text{co}(\Theta_0(\hat{P}_n)) = \Theta_0(\hat{P}_n)\}$. By Lemmas A.6 and B.5, $P(A_n^c) = o(1)$, and hence by Theorem 11.14 in Kosorok (2008) $P_{\eta/\sqrt{n}}(A_n^c) = o(1)$. Therefore, we obtain:

$$J_n(\theta_0) = \max\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n))), 0\} + o_{P_{\eta/\sqrt{n}}}(1) \tag{D.26}$$

since $J_n(\theta_0) = \max\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n))), 0\}$ whenever A_n occurs. Next, note that by Lemma B.8 in Kaido (2010), the map ψ is Hadamard differentiable at $\nu(\cdot, \Theta_0(P))$ with derivative $\dot{\psi} : \mathcal{C}(\mathbb{S}^{d_\theta}) \to \mathbf{R}$ pointwise given by:

$$\dot{\psi}(f) = -f(p_0) \ . \tag{D.27}$$

Moreover, the Hadamard differentiability of ψ together with Theorem 4.1 and Theorem 18.6 in Kosorok (2008) imply $\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n)))\}$ is an efficient estimator for $\psi(\nu(\cdot, \Theta_0(P)))$ 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:

$$\sqrt{n}\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n))) - \psi(\nu(\cdot, \Theta_0(P_{\eta/\sqrt{n}})))\} \xrightarrow{L_{\eta/\sqrt{n}}} -\mathbb{G}_0(p_0) \ . \tag{D.28}$$

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 A.2 and (D.27):

$$\frac{\partial}{\partial\eta}\psi(\nu(\cdot, \Theta_0(P_\eta)))\Big|_{\eta=0} = -2\int \lambda(p_0, P)'H(\theta_0)m(x, \theta_0)\dot{h}_0(x)h_0(x)d\mu(x) = 2\int \tilde{l}(x)\dot{h}_0(x)h_0(x)d\mu(x) \tag{D.29}$$

where $h_\eta \equiv \sqrt{dP_\eta/d\mu}$ and the final result holds by definition of $\tilde{l}(x)$ and $\int \dot{h}_0(x)h_0(x)d\mu(x) = 0$. Therefore,

$$\sqrt{n}\{\psi(\nu(\cdot, \Theta_0(\hat{P}_n))) - \psi(\nu(\cdot, \Theta_0(P)))\} \xrightarrow{L_{\eta/\sqrt{n}}} -\mathbb{G}_0(p_0) + \eta\int 2\tilde{l}(x)\dot{h}_0(x)h_0(x)d\mu(x) \ , \tag{D.30}$$

due to (D.28) and (D.29). Moreover, as shown in the proof of Theorem 5.3, $\hat{c}_{1-\alpha}(\theta_0) = c_{1-\alpha}(\theta_0) + o_p(1)$ when $X_i \sim P$ and therefore by Theorem 11.14 in Kosorok (2008) also when $X_i \sim P_{\eta/\sqrt{n}}$. Thus, exploiting result (D.26) we obtain:

$$\lim_{n\to\infty} P_{\eta/\sqrt{n}}(J_n(\theta_0) > \hat{c}_{1-\alpha}(\theta_0)) = \lim_{n\to\infty} P_{\eta/\sqrt{n}}(\max\{\psi(\nu(\cdot,\Theta_0(\hat{P}_n))),0\} > c_{1-\alpha}(\theta_0))$$
$$= \lim_{n\to\infty} P_{\eta/\sqrt{n}}(\psi(\nu(\cdot,\Theta_0(\hat{P}_n))) > c_{1-\alpha}(\theta_0)) = P(-\mathbb{G}_0(p_0) > c_{1-\alpha}(\theta_0) - 2\eta\int \tilde{l}(x)\dot{h}_0(x)h_0(x)d\mu(x)) \quad \text{(D.31)}$$

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 (D.30). Thus (D.31) verifies $\{\pi_n^*\}$ attains the bound in (34). Moreover, if $P_\eta \in \mathbf{H}(\theta_0)$, then by (D.29) we must have:

$$\int \tilde{l}(x)\dot{h}_0(x)h_0(x) \geq 0 \ . \quad \text{(D.32)}$$

Therefore, results (D.31) and (D.32) imply that $J_n(\theta_0)$ satisfies (33) as well.

We next establish the upper bound in (34) holds using arguments in the proof of Theorem 25.44 in van der Vaart (1999). Fix a $P_\eta \in \mathbf{H}(\theta_0)$ and $\bar{\eta} > 0$ for which we aim to show the bound, and pass to a subsequence $\{n_k\}_{k=1}^\infty$ with:

$$\limsup_{n\to\infty} \pi_n(P_{\bar{\eta}/\sqrt{n}}) = \lim_{k\to\infty} \pi_{n_k}(P_{\bar{\eta}/\sqrt{n_k}}) \ . \quad \text{(D.33)}$$

Further let $\tilde{s}(x) = 2\tilde{l}(x)h_0(x)$ and $\tilde{r}(x) = \tilde{s}(x) - \dot{h}_0(x)\langle\tilde{s},\dot{h}_0\rangle_{L^2_\mu}/\|\dot{h}_0\|^2_{L^2_\mu}$. Then, notice that by direct calculation we can obtain that $\tilde{s} \in \dot{\mathbf{S}}$, $\tilde{r} \in \dot{\mathbf{S}}$ and $\langle\tilde{r},\dot{h}_0\rangle_{L^2_\mu} = 0$. Moreover, also observe that by result (D.29) we have:

$$\langle\tilde{s},\dot{h}_0\rangle_{L^2_\mu} = \frac{\partial}{\partial\eta}\psi(\nu(\cdot,\Theta_0(P_\eta)))\Big|_{\eta=0} \ . \quad \text{(D.34)}$$

Proceeding as in the proof of Lemma A.16, we next build an augmented model by letting $s \equiv \sqrt{dP/d\mu}$, $\Psi : \mathbf{R} \to (0,\infty)$ be continuously differentiable, with $\Psi(0) = \Psi'(0) = 1$ and Ψ, Ψ' and Ψ'/Ψ bounded, and defining:

$$q^2_{\eta,\gamma}(x) \equiv b(\eta,\gamma)s^2(x)\Psi\Big(\frac{2}{s(x)}\{\eta\dot{h}_0(x)+\gamma\tilde{r}(x)\}\Big) \qquad b(\eta,\gamma) \equiv \Big[\int \Psi\Big(\frac{2}{s(x)}\{\eta\dot{h}_0(x)+\gamma\tilde{r}(x)\}\Big)dP(x)\Big]^{-1} \ . \quad \text{(D.35)}$$

For $Q_{\eta,\gamma}$ satisfying $q_{\eta,\gamma} = \sqrt{dQ_{\eta,\gamma}/d\mu}$, using Proposition 2.1.1 in Bickel et al. (1993) it is straightforward to verify that $(\eta,\gamma) \mapsto q_{\eta,\gamma}$ is then a quadratic mean differentiable model with $q_{0,0} = \sqrt{dP/d\mu}$. Moreover, Lemmas A.2, A.8, Corollary A.3, and $P \in \mathbf{P}$ satisfying Assumption 3.6(ii) imply that $Q_{\eta,\gamma} \in \mathbf{P}$ for all $(\eta,\gamma) \in N$ and N a suitably small neighborhood of $(0,0)$ in $\mathbf{R}^2$. By Theorems 12.2.3 and 13.4.1 in Lehmann and Romano (2005), it then follows that if $\|\tilde{r}\|^2_{L^2_\mu} \neq 0$, then there exists a further subsequence $\{n_{k_j}\}_{j=1}^\infty$ such that:

$$\lim_{j\to\infty} \pi_{n_{k_j}}(Q_{(\eta,\gamma)/\sqrt{n_{k_j}}}) = \pi(\eta,\gamma) \quad \text{(D.36)}$$

for all $(\eta,\gamma) \in N$, and where π is the power function of a test in a limit experiment that takes the form:

$$Z \sim N\Big(\begin{bmatrix}\eta\\ \gamma\end{bmatrix}, I_0^{-1}\Big) \qquad I_0 \equiv \begin{bmatrix} 4\|\dot{h}_0\|^2_{L^2_\mu} & 0 \\ 0 & 4\|\tilde{r}\|^2_{L^2_\mu}\end{bmatrix} \ . \quad \text{(D.37)}$$

Next we establish that the power function π corresponds to a test that controls size for the hypothesis:

$$H_0 : \eta\langle\dot{h}_0,\tilde{s}\rangle_{L^2_\mu} + \gamma\langle\tilde{r},\tilde{s}\rangle_{L^2_\mu} \leq 0 \qquad H_1 : \eta\langle\dot{h}_0,\tilde{s}\rangle_{L^2_\mu} + \gamma\langle\tilde{r},\tilde{s}\rangle_{L^2_\mu} > 0 \ . \quad \text{(D.38)}$$

Select any $(\eta_0,\gamma_0) \in \mathbf{R}^2$ such that $\eta_0\langle\dot{h}_0,\tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle\tilde{r},\tilde{s}\rangle_{L^2_\mu} < 0$ and define a path $t \mapsto \tilde{P}_t$ to be given by $\tilde{P}_t \equiv Q_{(-t\eta_0,-t\gamma_0)}$. Notice that $\tilde{P}_t \in \mathbf{P}$ for t small due to $Q_{(\eta,\gamma)} \in \mathbf{P}$ for all $(\eta,\gamma) \in N$. Then, as in (D.34):

$$\frac{\partial}{\partial t}\psi(\nu(\cdot,\Theta_0(\tilde{P}_t)))\Big|_{t=0} = -\{\eta_0\langle\dot{h}_0,\tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle\tilde{r},\tilde{s}\rangle_{L^2_\mu}\} > 0 \ , \quad \text{(D.39)}$$

and, in addition, since at $t = 0$, $\tilde{P}_0 = P$ we have $\psi(\nu(\cdot, \Theta_0(\tilde{P}_0))) = 0$ due to $\theta_0 \in \partial\Theta_0(P)$. Thus, from (D.39) we conclude $\tilde{P}_t \in \mathbf{H}(\theta_0)$ for t in a neighborhood of zero. Noting $Q_{(\eta_0,\gamma_0)/\sqrt{n}} = \tilde{P}_{-1/\sqrt{n}}$, it follows from (33) and (D.36):

$$\pi(\eta_0, \gamma_0) = \lim_{j\to\infty} \pi_{n_{k_j}}(Q_{(\eta_0,\gamma_0)/\sqrt{n_{k_j}}}) = \lim_{j\to\infty} \pi_{n_{k_j}}(\tilde{P}_{-1/\sqrt{n_{k_j}}}) \leq \limsup_{n\to\infty} \pi_n(\tilde{P}_{-1/\sqrt{n}}) \leq \alpha \; . \tag{D.40}$$

Since (D.40) holds for any (η_0, γ_0) such that $\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} < 0$, continuity of the power function π implies it also holds for any (η_0, γ_0) with $\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} = 0$. We conclude π corresponds to a test that controls size in (D.38). Therefore, Proposition 15.2 in van der Vaart (1999) and $\tilde{s}$ being in the linear span of $\dot{h}_0$ and $\tilde{r}$ yield:

$$\pi(\eta_0, \gamma_0) \leq 1 - \Phi\Big(z_{1-\alpha} - \frac{\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu}}{\sigma_0}\Big) \qquad \sigma_0^2 \equiv \frac{\langle \dot{h}_0, \tilde{s}\rangle^2_{L^2_\mu}}{4\|\dot{h}_0\|^2_{L^2_\mu}} + \frac{\langle \tilde{r}, \tilde{s}\rangle^2_{L^2_\mu}}{4\|\tilde{r}\|^2_{L^2_\mu}} = \frac{\|\tilde{s}\|^2_{L^2_\mu}}{4} \; , \tag{D.41}$$

for any (η_0, γ_0) such that $\eta_0\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + \gamma_0\langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} > 0$. Furthermore, since both $\eta \mapsto \sqrt{dP_\eta/d\mu}$ and $\eta \mapsto \sqrt{dQ_{\eta,0}/d\mu}$ are Fréchet differentiable in L^2_μ at $\eta = 0$ with derivative $\dot{h}_0$, we also have that for any $\bar{\eta} > 0$:

$$\limsup_{n\to\infty} \sqrt{n}\|h_{\bar{\eta}/\sqrt{n}} - q_{\bar{\eta}/\sqrt{n},0}\|_{L^2_\mu} \leq \limsup_{n\to\infty} \sqrt{n}\{\|h_{\bar{\eta}/\sqrt{n}} - h_0 - \frac{\bar{\eta}}{\sqrt{n}}\dot{h}_0\|_{L^2_\mu} + \|q_{\bar{\eta}/\sqrt{n},0} - h_0 - \frac{\bar{\eta}}{\sqrt{n}}\dot{h}_0\|_{L^2_\mu}\} = 0 \; . \tag{D.42}$$

Hence, by Theorem 13.1.4 in Lehmann and Romano (2005) $P^n_{\bar{\eta}/\sqrt{n}}$ and $Q^n_{\bar{\eta}/\sqrt{n},0}$ converge in total variation, and thus

$$\lim_{k\to\infty} \pi_{n_k}(P_{\bar{\eta}/\sqrt{n_k}}) = \lim_{k\to\infty} \pi_{n_k}(Q_{\bar{\eta}/\sqrt{n_k},0}) \; . \tag{D.43}$$

To conclude, observe that since $P_\eta \in \mathbf{H}(\theta_0)$, result (D.34) implies that $\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} \geq 0$. If $\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} > 0$, then $\bar{\eta} > 0$ and results (D.33), (D.36), (D.41) and (D.43) establish that:

$$\limsup_{n\to\infty} \pi_n(P_{\bar{\eta}/\sqrt{n}}) = \lim_{j\to\infty} \pi_{n_{k_j}}(Q_{(\bar{\eta},0)/\sqrt{n_{k_j}}}) = \pi(\bar{\eta}, 0) \leq 1 - \Phi\Big(z_{1-\alpha} - \frac{2\bar{\eta}E[\tilde{l}(X_i)\dot{h}_0(X_i)/h_0(X_i)]}{\sqrt{E[\mathbb{G}_0^2(p_0)]}}\Big) \tag{D.44}$$

where we have used $\sigma_0^2 = E[\mathbb{G}_0^2(p_0)]$, $\tilde{s}(x) = 2\tilde{l}(x)h_0(x)$ and $h_0^2 = dP/d\mu$. If on the other hand $\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} = 0$, then:

$$\limsup_{n\to\infty} \pi_n(P_{\bar{\eta}/\sqrt{n}}) = \lim_{j\to\infty} \pi_{n_{k_j}}(Q_{(\bar{\eta},0)/\sqrt{n_{k_j}}}) = \pi(\bar{\eta}, 0) \leq \alpha = 1 - \Phi\Big(z_{1-\alpha} - \frac{2\bar{\eta} \times 0}{\sqrt{E[\mathbb{G}_0^2(p_0)]}}\Big) \tag{D.45}$$

due to (D.33), (D.36), (D.43) together with $\bar{\eta}\langle \dot{h}_0, \tilde{s}\rangle_{L^2_\mu} + 0 \times \langle \tilde{r}, \tilde{s}\rangle_{L^2_\mu} = 0$ and π controlling size in (D.38). Recall we assumed $\|\tilde{r}\|_{L^2_\mu} \neq 0$ in obtaining (D.37), and hence the Theorem follows from (D.44) and (D.45) whenever $\|\tilde{r}\|_{L^2_\mu} \neq 0$. The case $\|\tilde{r}\|_{L^2_\mu} = 0$ follows from the arguments in (D.36)-(D.43) applied directly to P_η (rather than $Q_{\eta,\gamma}$). ■

Appendix E - Proof of Theorem 3.3

As in the proof of Theorem 3.2, we establish Theorem 3.3 by verifying the conditions of Theorem 5.2.1 in Bickel et al. (1993), which again requires us to: (i) Characterize the tangent space at P, and (ii) Show $Q \mapsto \nu(\cdot, \Theta_{0,I}(Q))$ is weakly pathwise differentiable at P. In this setting, however, both endeavors are simpler. Lemma E.1 employs Lemma A.16 to characterize the tangent space, while Lemma E.3 shows $Q \mapsto \nu(p, \Theta_{0,I}(Q))$ is weak-pathwise differentiable at P, and Lemma E.4 extends the result to show weak-pathwise differentiability of $Q \mapsto \nu(\cdot, \Theta_{0,I}(Q))$.

Subsequent to the proof of Theorem 3.3, we briefly discuss the connection between weak-pathwise differentiability in this setting, and in the moment inequalities model studied in Theorem 3.2.

Lemma E.1. *Let Assumption 3.7 hold, $P \in \mathbf{P_I}$, and $\mathbf{S_I} \equiv \{h \in L^2_\mu : h = \sqrt{dQ/d\mu}$ for some $Q \in \mathbf{P_I}\}$. Then the tangent space of $\mathbf{S_I}$ at $s = \sqrt{dP/d\mu}$ is $\dot{\mathbf{S}}_\mathbf{I} = \{h \in L^2_\mu : \int h(x)s(x)d\mu(x) = 0\}$.*

Proof: Let $P \in \mathbf{P_I}$ and $\xi(\int vz'dP(x))$ denote the smallest singular value of the matrix $\int vz'dP(x)$. Since $\mathcal{X}$ is compact by Assumption 3.7(i), it follows that vz' is bounded, and hence for any net $\{Q_\alpha\}_{\alpha\in\mathfrak{A}} \subset \mathbf{M}$ with $Q_\alpha \to P$:

$$\int vz'dQ_\alpha(x) \to \int vz'dP(x) \; . \tag{E.1}$$

Thus, since ξ is continuous under the Frobenius norm (Bhatia (1997), page 78) it follows from $P \in \mathbf{P_I}$ that there exists a neighborhood $N(P) \subseteq \mathbf{M}$ such that $\xi(\int vz'dQ(x)) > 0$ for all $Q \in N(P)$. We conclude that $\mathbf{P_I}$ is open in $\mathbf{M}_\mu \equiv \{Q \in \mathbf{M} : Q \ll \mu\}$ and the claim follows from Lemma A.16. ■

Lemma E.2. *Let Assumption 3.7 hold, and* $\mathbf{S_I} \equiv \{h \in L^2_\mu : h = \sqrt{dQ/d\mu}$ *for some* $Q \in \mathbf{P_I}\}$*. If* $\eta \mapsto h_\eta$ *is a curve in* $\mathbf{S_I}$ *and* $h_\eta = \sqrt{dP_\eta/d\mu}$*, then there is a neighborhood* $N \subset \mathbf{R}$ *of zero, such that for all* $\eta_0 \in N$*:*

$$\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}\Big|_{\eta=\eta_0} = -2\Sigma(P_{\eta_0})^{-1}\{\int vz'\dot{h}_{\eta_0}(x)h_{\eta_0}(x)d\mu(x)\}\Sigma(P_{\eta_0})^{-1} , \tag{E.2}$$

and in addition $\eta_0 \mapsto \frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}|_{\eta=\eta_0}$ *is continuous and* $\|\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}|_{\eta=\eta_0}\|_F$ *is uniformly bounded in* $\eta_0 \in N$*.*

Proof: Recall that if $\eta \mapsto U(\eta)$ is a square matrix valued function that is invertible at $\eta = \eta_0$, then $\frac{\partial}{\partial\eta}U(\eta)^{-1}|_{\eta=\eta_0} = -U(\eta_0)^{-1}\frac{\partial}{\partial\eta}U(\eta)|_{\eta=\eta_0}U(\eta_0)^{-1}$. Hence, since $P_\eta \in \mathbf{P_I}$ implies $\Sigma(P_\eta)$ is invertible, we obtain:

$$\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}\Big|_{\eta=\eta_0} = -\Sigma(P_{\eta_0})^{-1}\{\int 2vz'\dot{h}_{\eta_0}(x)h_{\eta_0}(x)d\mu(x)\}\Sigma(P_{\eta_0})^{-1} \tag{E.3}$$

by exploiting that vz' is bounded by Assumption 3.7(i), and arguing as in (A.43). Moreover, since $P_{\eta_0} \in \mathbf{P_I}$ by assumption, continuity of $\eta \mapsto \Sigma(P_\eta)^{-1}$ follows from (E.1) and $\|h_\eta - h_{\eta_0}\|_{L^2_\mu} = o(1)$ implying $P_\eta \to P_{\eta_0}$ in the τ-topology. Since vz' is uniformly bounded by Assumption 3.7(i), arguing as in (A.49) in turn implies that $\int 2vz'\dot{h}_{\eta_0}(x)h_{\eta_0}(x)d\mu(x)$ is continuous in η_0, and hence the continuity of $\eta_0 \mapsto \frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}|_{\eta=\eta_0}$ follows from (E.3). To conclude, note that $\|\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}|_{\eta=0}\|_F < \infty$ due to $\|\Sigma(P_0)^{-1}\|_F < \infty$, zv' being bounded, the Cauchy-Schwarz inequality, $\|h_0\|_{L^2_\mu} = 1$ and $\|\dot{h}_0\|_{L^2_\mu} < \infty$ because $\eta \mapsto h_\eta$ is Fréchet differentiable. Hence, since $\|\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}|_{\eta=0}\|_F$ is finite, continuity implies it must be uniformly bounded in a neighborhood of zero, and the Lemma follows. ■

Lemma E.3. *Let Assumption 3.7 hold, and* $\mathbf{S_I} \equiv \{h \in L^2_\mu : h = \sqrt{dQ/d\mu}$ *for some* $Q \in \mathbf{P_I}\}$*. If* $\eta \mapsto h_\eta$ *is a curve in* $\mathbf{S_I}$ *and* $h_\eta = \sqrt{dP_\eta/d\mu}$*, then there is a neighborhood* $N \subset \mathbf{R}$ *of zero, such that for all* $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N$*:*

$$\frac{\partial}{\partial\eta}\nu(p, \Theta_{0,I}(P_\eta))\Big|_{\eta=\eta_0} = 2\int\{\psi_\nu(p, x, P_{\eta_0}) - \psi_\Sigma(p, x, P_{\eta_0})\}\dot{h}_{\eta_0}(x)h_{\eta_0}(x)d\mu(x) , \tag{E.4}$$

where ψ_ν *and* ψ_Σ *are as defined in equations* (15) *and* (16) *respectively. In addition,* N *may be chosen so that* $(p, \eta_0) \mapsto \frac{\partial}{\partial\eta}\nu(p, \Theta_{0,I}(P_\eta))|_{\eta=\eta_0}$ *is continuous and uniformly bounded in* $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N$*.*

Proof: First note that since $P_\eta \in \mathbf{P_I}$ it follows that $\int vz'dP_\eta(x)$ is invertible, while $P_\eta \ll \mu$ and Assumption 3.7(ii) imply $P_\eta(Y_L \le Y_U) = 1$. Therefore, Proposition 2 in Bontemps et al. (2012) implies that:

$$\nu(p, \Theta_{0,I}(P_\eta)) = \int p'\Sigma(P_\eta)^{-1}v(y_L + 1\{p'\Sigma(P_\eta)^{-1}v > 0\}(y_U - y_L))dP_\eta(x) \tag{E.5}$$

provided $P_\eta(Y_L < Y_U) > 0$, while direct calculation shows (E.5) holds when $P_\eta(Y_L = Y_U) = 1$ since then $\Theta_{0,I}(P_\eta) = \{\Sigma(P_\eta)^{-1}\int vy_L dP_\eta(x)\}$. Let $\gamma_\eta(p, v) \equiv p'\Sigma(P_\eta)^{-1}v$ and note that if $(p_n, \eta_n) \to (p_0, \eta_0)$ with $p_0 \in \mathbb{S}^{d_\theta}$, then:

$$\mu((y_L, y_U, v, z) : \lim_{n\to\infty} 1\{\gamma_{\eta_n}(p_n, v) > 0\} = 1\{\gamma_{\eta_0}(p_0, v) > 0\}) = 1 \tag{E.6}$$

since $(p, \eta) \mapsto \gamma_\eta(p, v)$ is continuous, and $\mu((y_L, y_U, v, z) : p_0'\Sigma(P_{\eta_0})^{-1}v = 0) = 0$ by Assumption 3.7(iii). Moreover,

$$\lim_{n\to\infty}\sup_{p\in\mathbb{S}^{d_\theta}} |\int v^{(i)}(y_U - y_L)1\{\gamma_{\eta_n}(p, v) > 0\}(h^2_{\eta_n}(x) - h^2_{\eta_0}(x))d\mu(x)|$$
$$\le \sup_{x\in\mathcal{X}} 2\|x\|^2 \times \lim_{n\to\infty}\{\|h_{\eta_n} - h_{\eta_0}\|_{L^2_\mu} \times \|h_{\eta_n} + h_{\eta_0}\|_{L^2_\mu}\} = 0 , \tag{E.7}$$

for any $1 \le i \le d_Z$ by compactness of $\mathcal{X}$, the Cauchy-Schwarz inequality, $\|h_\eta\|_{L^2_\mu} = 1$ for all η, and $\eta \mapsto h_\eta$ being Fréchet differentiable. Hence, compactness of $\mathcal{X}$, result (E.6) and the dominated convergence theorem imply:

$$\lim_{n\to\infty}\int v(y_U - y_L)1\{\gamma_{\eta_n}(p, v) > 0\}h^2_{\eta_n}(x)d\mu(x) = \int v(y_U - y_L)1\{\gamma_{\eta_0}(p, v) > 0\}h^2_{\eta_0}(x)d\mu(x) . \tag{E.8}$$

Therefore, for any $p \in \mathbb{S}^{d_\theta}$ we can conclude from (E.8) and $\eta \mapsto \Sigma(P_\eta)^{-1}$ being differentiable by Lemma E.2 that:

$$\lim_{n\to\infty} \frac{1}{|\eta_n - \eta_0|} \int (\gamma_{\eta_n}(p,v) - \gamma_{\eta_0}(p,v))(y_U - y_L) 1\{\gamma_{\eta_n}(p,v) > 0\} h^2_{\eta_n}(x) d\mu(x)$$
$$= p'\{\frac{\partial}{\partial \eta}\Sigma(P_\eta)^{-1}\Big|_{\eta=\eta_0}\} \int v(y_U - y_L) 1\{\gamma_{\eta_0}(p,v) > 0\} h^2_{\eta_0}(x) d\mu(x) \ . \quad \text{(E.9)}$$

Next, note $\gamma_{\eta_0}(p,v)(y_U - y_L)$ is uniformly bounded by compactness of $\mathbb{S}^{d_\theta} \times \mathcal{X}$, and hence arguing as in (A.43):

$$\lim_{n\to\infty} \frac{1}{|\eta_n - \eta_0|} \int \gamma_{\eta_0}(p,v)(y_U - y_L) 1\{\gamma_{\eta_n}(p,v) > 0\} (h^2_{\eta_n}(x) - h^2_{\eta_0}(x) - 2(\eta_n - \eta_0)\dot{h}_{\eta_0}(x) h_{\eta_0}(x)) d\mu(x) = 0 \ . \quad \text{(E.10)}$$

Thus, results (E.6) and (E.10), compactness of $\mathcal{X}$, and the dominated convergence theorem yield:

$$\lim_{n\to\infty} \frac{1}{|\eta_n - \eta_0|} \int \gamma_{\eta_0}(p,v)(y_U - y_L) 1\{\gamma_{\eta_n}(p,v) > 0\} (h^2_{\eta_n}(x) - h^2_{\eta_0}(x)) d\mu(x)$$
$$= 2 \int \gamma_{\eta_0}(p,v)(y_U - y_L) 1\{\gamma_{\eta_0}(p,v) > 0\} \dot{h}_{\eta_0}(x) h_{\eta_0}(x) d\mu(x) \ . \quad \text{(E.11)}$$

In addition, Lemma E.2 and the mean value theorem imply that for some $\bar{\eta}_n(x)$ between η_n and η_0

$$\lim_{n\to\infty} |\int \gamma_{\eta_0}(p,v)(y_U - y_L)(1\{\gamma_{\eta_n}(p,v) > 0\} - 1\{\gamma_{\eta_0}(p,v) > 0\}) h^2_{\eta_0}(x) d\mu(x)|$$
$$= \lim_{n\to\infty} |\int \gamma_{\eta_0}(p,v)(y_U - y_L)(1\{\gamma_{\eta_0}(p,v) > (\eta_0 - \eta_n)\frac{\partial}{\partial\eta}\gamma_\eta(p,v)\Big|_{\eta=\bar{\eta}_n(x)}\} - 1\{\gamma_{\eta_0}(p,v) > 0\}) h^2_{\eta_0}(x) d\mu(x)|$$
$$\leq \lim_{n\to\infty} \int |\gamma_{\eta_0}(p,v)(y_U - y_L)| 1\{|\gamma_{\eta_0}(p,v)| \leq M|\eta_0 - \eta_n|\} h^2_{\eta_0}(x) d\mu(x) \ , \quad \text{(E.12)}$$

where the inequality holds for some $M > 0$ due to Lemma E.2 and compactness of $\mathbb{S}^{d_\theta} \times \mathcal{X}$ implying $\frac{\partial}{\partial\eta}\gamma_\eta(p,v)|_{\eta=\eta_0}$ is uniformly bounded for η_0 in a neighborhood of zero. Therefore, from (E.12) we conclude:

$$\lim_{n\to\infty} \frac{1}{|\eta_n - \eta_0|} |\int \gamma_{\eta_0}(p,v)(y_U - y_L)(1\{\gamma_{\eta_n}(p,v) > 0\} - 1\{\gamma_{\eta_0}(p,v) > 0\}) h^2_{\eta_0}(x) d\mu(x)|$$
$$\leq 2 \sup_{x\in\mathcal{X}} \|x\| \times \lim_{n\to\infty} M \int 1\{|\gamma_{\eta_0}(p,v)| \leq M|\eta_0 - \eta_n|\} h^2_{\eta_0}(x) d\mu(x) = 0 \ , \quad \text{(E.13)}$$

where the final equality results from the monotone convergence theorem, and $\mu((y_L, y_U, v, z) : p'\Sigma(P_{\eta_0})^{-1} v = 0) = 0$ by Assumption 3.7(iii) and $p'\Sigma(P_{\eta_0})^{-1} \neq 0$. Finally, combining results (E.9), (E.11) and (E.13) we can obtain:

$$\frac{\partial}{\partial\eta}\{\int \gamma_\eta(p,v)(y_U - y_L) 1\{\gamma_\eta(p,v) > 0\} h^2_\eta(x) d\mu(x)\}\Big|_{\eta=\eta_0}$$
$$= \int (p'\{\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}\Big|_{\eta=\eta_0}\} v h^2_{\eta_0}(x) + 2\gamma_{\eta_0}(p,v)\dot{h}_{\eta_0}(x) h_{\eta_0}(x))(y_U - y_L) 1\{\gamma_{\eta_0}(p,v) > 0\} d\mu(x) \ . \quad \text{(E.14)}$$

Similarly, Lemma E.2, compactness of $\mathcal{X}$ and arguing as in (E.9) and (E.11) allow us to establish that:

$$\frac{\partial}{\partial\eta}\{\int \gamma_\eta(p,v) y_L h^2_\eta(x) d\mu(x)\}\Big|_{\eta=\eta_0} = \int (p'\{\frac{\partial}{\partial\eta}\Sigma(P_\eta)^{-1}\Big|_{\eta=\eta_0}\} v h^2_{\eta_0}(x) + 2\gamma_{\eta_0}(p,v)\dot{h}_{\eta_0}(x) h_{\eta_0}(x)) y_L d\mu(x) \ . \quad \text{(E.15)}$$

Result (E.4) then follows from (E.14), (E.15) Lemma E.2, and the definitions of ψ_ν and ψ_Σ.

In order to establish continuity, let $(p_n, \eta_n) \to (p_0, \eta_0) \in \mathbb{S}^{d_\theta} \times N$. Results (E.6) and (E.7) then imply that:

$$\lim_{n\to\infty} \int v(y_U - y_L) 1\{\gamma_{\eta_n}(p_n,v) > 0\} h^2_{\eta_n}(x) d\mu(x) = \int v(y_U - y_L) 1\{\gamma_{\eta_0}(p_0,v) > 0\} h^2_{\eta_0}(x) d\mu(x) \quad \text{(E.16)}$$

by the dominated convergence theorem. Next, note that by compactness of $\mathcal{X}$ and the Cauchy-Schwarz inequality:

$$\lim_{n\to\infty} |\int v^{(i)}(y_U - y_L) 1\{\gamma_{\eta_n}(p_n,v) > 0\} (\dot{h}_{\eta_n}(x) h_{\eta_n}(x) - \dot{h}_{\eta_0}(x) h_{\eta_0}(x)) d\mu(x)|$$
$$\leq 2 \sup_{x\in\mathcal{X}} \|x\|^2 \times \lim_{n\to\infty} \{\|\dot{h}_{\eta_n} - \dot{h}_{\eta_0}\|_{L^2_\mu} \|h_{\eta_n}\|_{L^2_\mu} + \|h_{\eta_n} - h_{\eta_0}\|_{L^2_\mu} \|\dot{h}_{\eta_0}\|_{L^2_\mu}\} = 0 \quad \text{(E.17)}$$

since $\|h_\eta\|_{L^2_\mu} = 1$ for all η and $\eta \mapsto h_\eta$ is continuously Fréchet differentiable. Hence, we can conclude that:

$$\lim_{n\to\infty} \int v(y_U - y_L)1\{\gamma_{\eta_n}(p_n, v) > 0\}\dot{h}_{\eta_n}(x)h_{\eta_n}(x)d\mu(x) = \int v(y_U - y_L)1\{\gamma_{\eta_0}(p_0, v) > 0\}\dot{h}_{\eta_0}(x)h_{\eta_0}(x)d\mu(x) \quad \text{(E.18)}$$

by (E.6) and the dominated convergence theorem. Therefore, (E.14), (E.16), (E.18) and Lemma E.2 yield:

$$\lim_{n\to\infty} \frac{\partial}{\partial\eta}\{\int \gamma_\eta(p_n, v)(y_U - y_L)1\{\gamma_\eta(p_n, v) > 0\}h^2_\eta(x)d\mu(x)\}\Big|_{\eta=\eta_n}$$
$$= \frac{\partial}{\partial\eta}\{\int \gamma_\eta(p_0, v)(y_U - y_L)1\{\gamma_\eta(p_0, v) > 0\}h^2_\eta(x)d\mu(x)\}\Big|_{\eta=\eta_0} . \quad \text{(E.19)}$$

Similarly, employing the same arguments as in (E.16) and (E.18) together with result (E.15) it is possible to show:

$$\lim_{n\to\infty} \frac{\partial}{\partial\eta}\{\int \gamma_\eta(p_n, v)y_L h^2_\eta(x)d\mu(x)\}\Big|_{\eta=\eta_n} = \frac{\partial}{\partial\eta}\{\int \gamma_\eta(p_0, v)y_L h^2_\eta(x)d\mu(x)\}\Big|_{\eta=\eta_0} . \quad \text{(E.20)}$$

Thus, continuity of $(p, \eta_0) \mapsto \frac{\partial}{\partial\eta}\nu(p, \Theta_{0,I}(P_\eta))|_{\eta=\eta_0}$ follows from (E.5), (E.19) and (E.20). Finally, note that since $\eta \mapsto h_\eta$ is continuously Fréchet differentiable, we may choose the neighborhood $N \subseteq \mathbf{R}$ so that $\|\dot{h}_\eta\|_{L^2_\mu}$ is uniformly bounded in $\eta \in N$. The Cauchy-Schwarz inequality then implies $|\int \dot{h}_\eta(x)h_\eta(x)d\mu(x)| \leq \|\dot{h}_\eta\|_{L^2_\mu}\|h_\eta\|_{L^2_\mu} < \infty$ uniformly in $\eta \in N$. Therefore, compactness of $\mathcal{X} \times \mathbb{S}^{d_\theta}$, Lemma E.2, and results (E.5), (E.14) and (E.15) imply $\frac{\partial}{\partial\eta}\nu(p, \Theta_{0,I}(P_\eta))|_{\eta=\eta_0}$ is uniformly bounded in $(p, \eta_0) \in \mathbb{S}^{d_\theta} \times N$, and the Lemma follows. ■

Lemma E.4. *Let Assumption 3.7 hold, and $\rho_I : \mathbf{P_I} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ be given by $\rho_I(P) \equiv \nu(\cdot, \Theta_{0,I}(P))$. Then ρ_I is weakly pathwise differentiable at any $P \in \mathbf{P_I}$, and for $s \equiv \sqrt{dP/d\mu}$ the derivative $\dot{\rho}_I : \dot{\mathbf{S}}_\mathbf{I} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ satisfies:*

$$\dot{\rho}_I(\dot{h}_0)(p) = 2\int \{\psi_\nu(p, x, P) - \psi_\Sigma(p, x, P)\}\dot{h}_0(x)h_0(x)d\mu(x) ,$$

where ψ_ν and ψ_Σ are as defined in equations (15) *and* (16) *respectively.*

Proof: We first note that Lemma E.3 implies $\dot{\rho}_I(\dot{h}_0) \in \mathcal{C}(\mathbb{S}^{d_\theta})$ for any $\dot{h}_0 \in \dot{\mathbf{S}}_\mathbf{I}$. In addition, $\dot{\rho}_I$ is linear by inspection, while $\psi_\nu(p, x, P)$ and $\psi_\Sigma(p, x, P)$ being uniformly bounded in $(p, x) \in \mathbb{S}^{d_\theta} \times \mathcal{X}$ by Assumption 3.7(i) imply:

$$\sup_{\|\dot{h}_0\|_{L^2_\mu}=1} \|\dot{\rho}_I(\dot{h}_0)\|_\infty \leq \sup_{(p,x)\in\mathbb{S}^{d_\theta}\times\mathcal{X}} 2\{|\psi_\nu(p, x, P)| + |\psi_\Sigma(p, x, P)|\} \times \sup_{\|\dot{h}_0\|_{L^2_\mu}=1} \{\|\dot{h}_0\|_{L^2_\mu} \times \|h_0\|_{L^2_\mu}\} < \infty , \quad \text{(E.21)}$$

and hence $\dot{\rho}_I$ is continuous as well. Moreover, for any finite Borel measure B on $\mathbb{S}^{d_\theta}$ and curve $\eta \mapsto P_\eta \in \mathbf{P_I}$ with $h_0 = s$, the mean value and dominated convergence theorems together with Lemma E.3 yield:

$$\lim_{\eta_0\to 0} \int \{\frac{\nu(p, \Theta_{0,I}(P_{\eta_0})) - \nu(p, \Theta_{0,I}(P))}{\eta_0} - \dot{\rho}_I(\dot{h}_0)(p)\}dB(p) = 0 , \quad \text{(E.22)}$$

(see also (A.57)). Result (E.22) verifies $\dot{\rho}_I$ is the weak derivative of ρ_I and the Lemma follows. ■

Proof of Theorem 3.3: As in the proof of Theorem 3.2, we let $\mathbf{B} \equiv \mathcal{C}(\mathbb{S}^{d_\theta})$ and $\mathbf{B}^*$ denote the set of finite Borel measures on $\mathbb{S}^{d_\theta}$, which is the dual of $\mathbf{B}$ by Corollary 14.15 in Aliprantis and Border (2006). Let $\rho_I : \mathbf{P_I} \to \mathbf{B}$ be given by $\rho_I(P) \equiv \nu(\cdot, \Theta_{0,I}(P))$, which has weak derivative $\dot{\rho}_I$ by Lemma E.4. For any $B \in \mathbf{B}^*$ then define:

$$\dot{\rho}_I^T(B)(x) \equiv 2\int_{\mathbb{S}^{d_\theta}} \{\psi(x, p, P) - E[\psi(X_i, p, P)]\}s(x)dB(p) , \quad \text{(E.23)}$$

where $s \equiv \sqrt{dP/d\mu}$, and the measurability of the integrand can be established arguing as in (A.59). In what follows, we aim to show $\dot{\rho}_I^T : \mathbf{B}^* \to \dot{\mathbf{S}}_\mathbf{I}$ is the adjoint of $\dot{\rho}_I : \dot{\mathbf{S}}_\mathbf{I} \to \mathbf{B}$. To this end, note $\dot{\rho}_I^T(B) \in L^2_\mu$ for any $B \in \mathbf{B}^*$ since $\psi(p, x, P) = \psi_\nu(p, x, P) - \psi_\Sigma(p, x, P)$ is uniformly bounded in $(p, x) \in \mathbb{S}^{d_\theta} \times \mathcal{X}$ as argued in (E.21). Moreover,

$$\int_\mathcal{X} \dot{\rho}_I^T(B)(x)s(x)d\mu(x) = 2\int_{\mathbb{S}^{d_\theta}}\int_\mathcal{X} \{\psi(x, p, P) - E[\psi(X_i, p, P)]\}dP(x)dB(p) = 0 , \quad \text{(E.24)}$$

by exchanging the order of integration and exploiting that $s^2 = dP/d\mu$. Hence, Lemma E.1 and (E.24) verify that $\dot{\rho}_I^T(B) \in \dot{\mathbf{S}}_\mathbf{I}$ for any $B \in \mathbf{B}^*$. Finally, for any $\dot{h}_0 \in \dot{\mathbf{S}}_\mathbf{I}$ and $B \in \mathbf{B}^*$ we can use that $\int \dot{h}_0(x)s(x)d\mu(x) = 0$ by Lemma

E.1, exchange the order of integration and exploit Lemma E.4 to obtain that:

$$\int_{\mathcal{X}} \dot{\rho}_I^T(B)(x)\dot{h}_0(x)d\mu(x) = \int_{\mathbb{S}^{d_\theta}} \int_{\mathcal{X}} \psi(x,p,P)\dot{h}_0(x)s(x)d\mu(x)dB(p) = \int_{\mathbb{S}^{d_\theta}} \dot{\rho}_I(\dot{h}_0)(p)dB(p) \ . \tag{E.25}$$

From result (E.25) we conclude $\dot{\rho}_I^T : \mathbf{B}^* \to \dot{\mathbf{S}}_{\mathbf{I}}$ is indeed the adjoint of $\dot{\rho}_I : \dot{\mathbf{S}}_{\mathbf{I}} \to \mathbf{B}$, and the Theorem then follows from Theorem 5.2.1 in Bickel et al. (1993). ■

The principal challenge in establishing Theorem 3.3 is in verifying weak-pathwise differentiability of the support function of the identified set. Differentiability of the support function in particular implies that the scalar valued parameter $Q \mapsto \nu(p_0, \Theta_{0,I}(Q))$ must be differentiable at every $p_0 \in \mathbb{S}^{d_\theta}$, which by (14) is equivalent to

$$\nu(p_0, \Theta_{0,I}(P_\eta)) = \int p_0'\Sigma(P_\eta)^{-1}v(y_L + 1\{p_0'\Sigma(P_\eta)^{-1}v > 0\}(y_U - y_L))dP_\eta(x) \tag{E.26}$$

being differentiable in η for any parametric submodel $\eta \mapsto P_\eta$. Inspecting (E.26), however, reveals that nondifferentiability at $\eta = 0$ may occur if $P(p_0'\Sigma(P)^{-1}V = 0) > 0$ – a situation that is ruled out by Assumption 3.7(iii). Interestingly, when V is a discrete random vector, the identified set $\Theta_{0,I}(P)$ has "flat" or "exposed" faces, and the $p_0 \in \mathbb{S}^{d_\theta}$ such that $P(p_0'\Sigma(P)^{-1}V = 0) > 0$ are precisely the $p_0 \in \mathbb{S}^{d_\theta}$ that are orthogonal to these flat faces; see Bontemps et al. (2012). In close connection to Remark 3.2, it is then possible to show $Q \mapsto \nu(p_0, \Theta_{0,I}(Q))$ is not pathwise weak-differentiable at any such p_0 by constructing a path $\eta \mapsto P_\eta$ that alters the slope of the exposed face.

Example E.1. Suppose $Z = V = (1, W)'$, $W \in \{-1, 0, 1\}$, and $Y_L, Y_U \in \mathcal{Y} \subset \mathbf{R}$ with $\mathcal{Y}$ compact. Further let $X = (Y_L, Y_U, V')'$, $\mathcal{X} = \mathcal{Y} \times \mathcal{Y} \times \{1\} \times \{-1, 0, 1\}$, and $\mu \in \mathbf{M}$ satisfy Assumption 3.7(ii). The set of $\theta = (\alpha, \beta)$ with

$$E[\tilde{Y} - \alpha - W\beta] = 0 \qquad\qquad E[W(\tilde{Y} - \alpha - W\beta)] = 0 \tag{E.27}$$

for some $\tilde{Y}$ satisfying $Y_L \leq \tilde{Y} \leq Y_U$, then constitutes the identified set under P. Further suppose P is such that

$$P(W = -1) = P(W = 0) = P(W = 1) = \frac{1}{3} \ , \tag{E.28}$$

for $a \in \{-1, 0, 1\}$ and $\ell \in \{L, U\}$ define $E_P[Y_\ell | W = a] \equiv \int y_\ell 1\{w = a\}dP(x)/P(W = a)$, and for simplicity let

$$E_P[Y_L|W = 0] = E[Y_U|W = 0] = 0 \ . \tag{E.29}$$

Let us consider a submodel satisfying $E_{P_\eta}[Y_\ell|W = a] = E_P[Y_\ell|W = a]$ for all $a \in \{-1, 0, 1\}$ and $\ell \in \{L, U\}$, and

$$P_\eta(W = -1) = \frac{1}{3}(1 - \eta) \qquad P_\eta(W = 0) = \frac{1}{3}(1 + 2\eta) \qquad P_\eta(W = 1) = \frac{1}{3}(1 - \eta) \ . \tag{E.30}$$

Along the submodel $\eta \mapsto P_\eta$, we can then obtain by direct calculation that the identified set at P_η is given by

$$\Theta_0(P_\eta) = \Big\{\theta \in \mathbf{R}^2 : \begin{array}{l} \text{(i) } E_{P_\eta}[Y_L|W = -1] \leq \frac{3}{2}\frac{\alpha}{1-\eta} - \beta \leq E_{P_\eta}[Y_U|W = -1] \\ \text{(ii) } E_{P_\eta}[Y_L|W = 1] \leq \frac{3}{2}\frac{\alpha}{1-\eta} + \beta \leq E_{P_\eta}[Y_U|W = 1] \end{array} \Big\} \ . \tag{E.31}$$

Thus, $\Theta_0(P_\eta)$ is a parallelogram with the slope of exposed faces depending on η. As in Remark 3.2, $\eta \mapsto \nu(p_0, \Theta_0(P_\eta))$ is not differentiable at $\eta = 0$ for an appropriate choice of p_0. For instance, for $p_0 = (\frac{3}{\sqrt{13}}, \frac{2}{\sqrt{13}})$ we obtain by (E.26)

$$\nu(p_0, \Theta_0(P_\eta)) = \frac{2-\eta}{\sqrt{13}}E[Y_U|W = 1] - \frac{\eta}{\sqrt{13}}(E[Y_L|W = -1] + E[Y_U - Y_L|W = -1]1\{\eta < 0\}) \ , \tag{E.32}$$

which is not differentiable at $\eta = 0$ if $E[Y_U - Y_L|W = -1] \neq 0$. Thus, $\eta \mapsto \nu(p_0, \Theta_0(P_\eta))$ is not differentiable at $\eta = 0$ precisely at a p_0 that is orthogonal to one of the exposed faces of the identified set $\Theta_0(P)$. ■

Appendix F - Discussion of Examples 2.1, 2.2, 2.3 and 2.4

In this Appendix we revisit Examples 2.1, 2.2, 2.3 and 2.4 from the main text. We map each example into our general framework, and examine Assumptions 3.2, 3.3, 3.4, 3.5 and 3.6 in their context.

Example 2.1 (Interval Censored Outcome)

In this example, $X = (Y_L, Y_U, Z')'$ and we let $\mathcal{Y} \subseteq \mathbf{R}$, $\mathcal{Z} = \{z_1, \ldots, z_K\}$ with $K < \infty$ and $\mathcal{X} = \mathcal{Y} \times \mathcal{Y} \times \mathcal{Z}$. For $\mathbf{M}$ the set of Borel probability measures on $\mathcal{X}$ and any $Q \in \mathbf{M}$ such that $Q(Z = z_k) > 0$, then denote for $\ell \in \{L, U\}$

$$E_Q[Y_\ell | Z = z_k] \equiv \frac{\int y_\ell 1\{Z = z_k\} dQ(x)}{\int 1\{Z = z_k\} dQ(x)} . \tag{F.1}$$

For a parameter space $\Theta \subseteq \mathbf{R}^{d_\theta}$ and any $Q \in \mathbf{M}$, then recall that in this example the identified set under Q is

$$\Theta_0(Q) \equiv \{\theta \in \Theta : E_Q[Y_L | Z = z_k] \leq z_k'\theta \leq E_Q[Y_U | Z = z_k] \text{ for all } 1 \leq k \leq K\} . \tag{F.2}$$

To map this setting into the framework of (2) and (3), we let $1_{\mathcal{Z}}(z) \equiv (1\{z = z_1\}, \ldots, 1\{z = z_K\})'$ and $m_S(x, \theta) = (y_L 1_{\mathcal{Z}}(z)', y_U 1_{\mathcal{Z}}(z)', 1_{\mathcal{Z}}(z)')'$ for all $\theta \in \Theta$. Then define $F_S : \mathbf{R}^{3K} \to \mathbf{R}^{2K}$ to be pointwise given by

$$F_S^{(i)}(v) \equiv \begin{cases} \frac{v^{(i)}}{v^{(2K+i)}}, & i = 1, \ldots, K \\ -\frac{v^{(i)}}{v^{(K+i)}}, & i = K+1, \ldots, 2K . \end{cases} \tag{F.3}$$

If $Q \in \mathbf{M}$ satisfies $Q(Z = z_k) > 0$ for some $1 \leq k \leq K$, then (F.3) implies $F_S^{(k)}(\int m_S(x, \theta) dQ(x)) = E_Q[Y_L | Z = z_k]$ and $F_S^{(2k)}(\int m_S(x, \theta) dQ(x)) = -E_Q[Y_U | Z = z_k]$. Hence, setting $A = (-z_1, \ldots, -z_K, z_1, \ldots, z_K)'$ we obtain

$$\Theta_0(Q) = \{\theta \in \Theta : A\theta + F_S(\int m_S(x, \theta) dQ(x)) \leq 0\} . \tag{F.4}$$

The following more primitive assumptions suffice for verifying Assumptions 3.2-3.6 in this example.

Assumption F.1. *(i) $\mathcal{Y}$ is compact; (ii) $\Theta \equiv \{\theta \in \mathbf{R}^{d_\theta} : \|\theta\|^2 \leq B_0\}$ with $B_0 < \infty$ satisfying $C_0 B_0 > K\{\sup_{y \in \mathcal{Y}} y^2\}$ where $C_0 \equiv \inf_{p \in \mathbb{S}^{d_\theta}} \sum_k \langle p, z_k \rangle^2$; (iii) $K \geq d_\theta$; (iv) Any subset $\mathcal{C} \subseteq \mathcal{Z}$ with $\#\mathcal{C} \leq d_\theta$ is linearly independent.*

Assumption F.2. *(i) For some $\theta_0 \in \mathbf{R}^{d_\theta}$, $E_P[Y_L | Z = z_k] \leq z_k'\theta_0 \leq E_P[Y_U | Z = z_k]$ for all $1 \leq k \leq K$; (ii) $P(Z = z_k) > 0$ and $E_P[Y_L - Y_U | Z = z_k] < 0$ for all $1 \leq k \leq K$; (iii) $\#\mathcal{A}(\theta, P) \leq d_\theta$ for all $\theta \in \Theta_0(P)$.*

Assumption F.1(i) imposes that Y_L and Y_U have compact support, which we require to verify Assumption 3.4(i). Assumption F.1(ii) defines Θ to be a ball of radius $\sqrt{B_0}$, where B_0 is chosen to ensure that $\Theta_0(P) \subset \Theta^o$ as required by Assumption 3.6(i). Assumptions F.1(iii)-(iv) impose a linear independence restrictions on the support points of Z, which together guarantee that $\Theta_0(P)$ is bounded. Assumption F.2 contains the main requirements on P. In particular, Assumption F.2(i), which holds if the model is properly specified, guarantees that $\Theta_0(P) \neq \emptyset$. The requirement $E_P[Y_L - Y_U | Z = z_k] < 0$ ensures that there is no $\theta \in \Theta_0(P)$ such that $E_P[Y_L | Z = z_k] = z_k'\theta = E_P[Y_U | Z = z_k]$, which would violate Assumption 3.6(iv). Finally, Assumption F.2(iii) requires that the number of binding constraints at each $\theta \in \Theta_0(P)$ be less than or equal to d_θ, and together with Assumption F.1(iv) imply Assumption 3.6(iv). We note that if $K = d_\theta$, then Assumption F.1(iv) and $E_P[Y_L - Y_U | Z = z_k] < 0$ imply Assumption F.2(iii) is automatically satisfied. In general, however, Assumption F.2(iii) imposes additional requirements on P.

Proposition F.1. *In Example 2.1, Assumptions F.1 and F.2 imply Assumptions 3.2-3.6.*

Proof: Assumption 3.2(i) is implied by Assumption F.1(ii). Further note that since the $2K \times d_\theta$ matrix A is known, Assumption 3.3(i) holds. Moreover, since $\mathcal{Y}$ is compact by Assumption F.1(i), $m_S(x, \theta) = (y_L 1_{\mathcal{Z}}(z)', y_U 1_{\mathcal{Z}}(z)', 1_{\mathcal{Z}}(z)')'$ is uniformly bounded in $\mathcal{X} \times \Theta$ and hence $m(x, \theta) = (m_S(x, \theta)', \theta' A')'$ and Θ being compact by Assumption F.1(ii) verify Assumption 3.4(i). In addition, given the definition of $m_S(x, \theta)$, Assumptions 3.4(ii)-(iii) directly follow from:

$$\nabla_\theta m(x, \theta) = \nabla_\theta \left[\begin{array}{c} m_S(x, \theta) \\ A\theta \end{array} \right] = \left[\begin{array}{c} 0 \\ A \end{array} \right] . \tag{F.5}$$

In order to verify Assumption 3.5, set $0 < \epsilon_0 < \inf_k P(Z = z_k)$, which is possible by Assumption F.2(ii), and $M_0 > 0$ so that $\max\{\sup_{y \in \mathcal{Y}} |y|, B_0 \sup_{z \in \mathcal{Z}} \|z\|\} < M_0 < \infty$, which is possible by compactness of $\mathcal{Y}$. Then defining

$$V_0 \equiv (-M_0, M_0)^{2K} \times (\epsilon_0, 1)^K \times (-M_0, M_0)^{2K} , \tag{F.6}$$

and noting $F(v)$ is differentiable unless $v^{(i)} = 0$ for some $2K + 1 \leq i \leq 3K$, it follows that Assumption 3.5(i) holds. Moreover, since ∇F is continuous on the closure of V_0 and V_0 is precompact, Assumption 3.5(ii) holds as well.

We next verify that P satisfies Assumption 3.6. First observe that Assumption F.2(i) implies $\theta_0 \in \Theta_0(P)$ and hence $\Theta_0(P) \neq \emptyset$. Next, also note that if $\theta \in \Theta_0(P)$, then (F.2) implies that for any $1 \leq k \leq K$:

$$|z_k'\theta| \leq \max\{|E[Y_L|Z = z_k]|, |E[Y_U|Z = z_k]|\} \leq \sup_{y \in \mathcal{Y}} |y| \ . \tag{F.7}$$

Furthermore, Assumptions F.1(iii)-(iv) imply $\mathbf{R}^{d_\theta} = \text{span}\{z_1, \ldots, z_K\}$, and hence $C_0 = \inf_{p \in \mathbb{S}^{d_\theta}} \sum_k \langle p, z_k \rangle^2 > 0$ by compactness of $\mathbb{S}^{d_\theta}$. Therefore, since $\theta/\|\theta\| \in \mathbb{S}^{d_\theta}$, we obtain from (F.7) that for any $\theta \in \Theta_0(P)$

$$\|\theta\|^2 C_0 \leq \|\theta\|^2 \sum_{k=1}^{K} \langle z_k, \frac{\theta}{\|\theta\|} \rangle^2 \leq K \sup_{y \in \mathcal{Y}} y^2 \ . \tag{F.8}$$

It then follows from Assumption F.1(ii) that if $\theta \in \Theta_0(P)$, then $\|\theta\|^2 < B_0$ and hence $\Theta_0(P) \subseteq \Theta^o$. However, since $\Theta_0(P)$ is closed, we must have $\Theta_0(P) \subset \Theta^o$, which verifies Assumption 3.6(i).

Since $m_S(x, \theta) = (y_L 1_{\mathcal{Z}}(z)', y_U 1_{\mathcal{Z}}(z)', 1_{\mathcal{Z}}(z)')'$ does not depend on θ, it follows that $\mathcal{S}_i = \emptyset$ for all $1 \leq i \leq 2K$ (see (4)), and hence Assumption 3.6(ii) actually holds for all $Q \in \mathbf{M}$. In turn, by definitions of ϵ_0 and M_0 we also have $\int m(x, \theta) dP(x) \in V_0$ for all $\theta \in \Theta$ and thus Assumption 3.6(iii) holds as well. Finally, note that

$$\nabla F^{(i)}\Big(\int m(x, \theta) dP(x)\Big) \int \nabla_\theta m(x, \theta) dP(x) = \begin{cases} -z_i & \text{if } 1 \leq i \leq K \\ +z_i & \text{if } K + 1 \leq i \leq 2K \end{cases} . \tag{F.9}$$

For notational simplicity, let $\mathcal{P}(\theta) = \{\nabla F^{(i)}(\int m(x, \theta) dP(x)) \int \nabla_\theta m(x, \theta) dP(x)\}_{i \in \mathcal{A}(\theta, P)}$. Then note that since $E_P[Y_L - Y_U|Z = z_k] < 0$, it follows for $1 \leq i \leq K$ that if $i \in \mathcal{A}(\theta, P)$ then $K + i \notin \mathcal{A}(\theta, P)$ – or equivalently, if $-z_i \in \mathcal{P}(\theta)$, then $z_i \notin \mathcal{P}(\theta)$. Assumptions F.1(iv) and F.2(iii) then imply the elements of $\mathcal{P}(\theta)$ are linearly independent for all $\theta \in \Theta_0(P)$, which verifies Assumption 3.6(iv). ■

Example 2.2 (Discrete Choice)

The structure of this example is identical to that of Example 2.1, though the notation is substantially more cumbersome. In this example $X = (Y, Z^{*\prime})'$, and we let $\mathcal{Y} \subseteq \mathbf{R}$. Also recall Z^* is assumed to have finite support $\mathcal{Z} = \{z_1, \ldots, z_K\}$ with $K < \infty$. Set $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$, and let $\mathbf{M}$ denote the set of Borel measures on $\mathcal{X}$. For notational convenience, we also define $\Delta(y, z_j, z_k) \equiv \psi(y, z_j) - \psi(y, z_k)$ and the set $\mathcal{V}$ to be given by

$$\mathcal{V} \equiv \{z_1 - z_2, \ldots, z_1 - z_K, z_2 - z_3, \ldots, z_2 - z_K, \ldots, z_{K-1} - z_K\} \ . \tag{F.10}$$

For any $Q \in \mathbf{M}$ such that $Q(Z^* = z_k) > 0$ let $E_Q[\Delta(Y, z_j, z_k)|Z^* = z_k] \equiv \int \Delta(y, z_j, z_k) dQ(x) / \int 1\{z^* = z_k\} dQ(x)$ (as in (F.1)), and note that for a parameter space Θ, the identified set under $Q \in \mathbf{M}$ in this example is then

$$\Theta_0(Q) \equiv \{\theta \in \Theta : E_Q[\Delta(Y, z_j, z_k)|Z^* = z_k] + (z_j - z_k)'\theta \leq 0 \text{ for all } z_j \neq z_k\} \ . \tag{F.11}$$

In order to identify (F.11) with the framework of (2) and (3), for each $1 \leq k \leq K$ let $v_k(y, z^*) \in \mathbf{R}^{K-1}$ satisfy

$$v_k^{(j)}(y, z^*) = \begin{cases} \Delta(y, z_j, z_k) 1\{z^* = z_k\}, & 1 \leq j < k \\ \Delta(y, z_{j+1}, z_k) 1\{z^* = z_k\}, & k \leq j \leq K - 1 \ . \end{cases} \tag{F.12}$$

Then let $v(y, z^*) = (v_1(y, z^*)', \ldots, v_K(y, z^*)')'$, $1_{\mathcal{Z}}(z^*) \equiv (1\{z^* = z_1\}, \ldots, 1\{z^* = z_K\})'$ and set $m_S(x, \theta) \in \mathbf{R}^{K^2}$ to be given by $m_S(x, \theta) = (v(y, z^*)', 1_{\mathcal{Z}}(z^*)')'$. We can then define $F_S : \mathbf{R}^{K^2} \to \mathbf{R}^{K(K-1)}$ to be pointwise given by

$$F_S^{(i)}(v) = \frac{v^{(i)}}{v^{(K(K-1) + \lceil \frac{i}{K-1} \rceil)}} \qquad i = 1, \ldots, K(K - 1) \ , \tag{F.13}$$

where $\lceil c \rceil$ denotes the smallest integer k such that $k \geq c$. Given these definitions, if $Q \in \mathbf{M}$ is such that $Q(Z^* = z_k) > 0$ and $(K - 1)(k - 1) + 1 \leq i \leq (K - 1)k$ then $F_S^{(i)}(\int m_S(x, \theta) dQ(x)) = E_Q[\Delta(Y, z_j, z_k)|Z^* = z_k]$ for some

$j \neq k$. Moreover, by setting $A = ((z_1 - z_2)', \ldots, (z_1 - z_K)', \ldots, (z_K - z_1)', \ldots, (z_K - z_{K-1})')'$ we obtain

$$\Theta_0(Q) = \{\theta \in \Theta : A\theta + F_S(\int m_S(x,\theta)dQ(x)) \leq 0\} . \tag{F.14}$$

Given the identical structure of Examples 2.1 and 2.2, we can derive sufficient conditions for Assumptions 3.2-3.6 by recasting Assumptions F.1 and F.2 in the present context. A formal proof that Assumptions F.3 and F.4 imply 3.2-3.6 can be obtained by arguments identical to those of Proposition F.1 and is therefore omitted.

Assumption F.3. *(i) $\psi : \mathcal{Y} \times \mathcal{Z} \to \mathbf{R}$ is bounded; (ii) $\Theta \equiv \{\theta \in \mathbf{R}^{d_\theta} : \|\theta\|^2 \leq B_0\}$ with $B_0 < \infty$ satisfying $2C_0B_0 > K(K-1)\{\sup_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}(\psi(y,z))^2\}$ where $C_0 \equiv \inf_{p\in\mathbb{S}^{d_\theta}} \sum_{v\in\mathcal{V}}\langle p, v\rangle^2$; (iii) $K(K-1) \geq 2d_\theta$; (iv) Any subset $\mathcal{C} \subseteq \mathcal{V}$ satisfying $\#\mathcal{C} \leq d_\theta$ is linearly independent.*

Assumption F.4. *(i) For some $\theta_0 \in \mathbf{R}^{d_\theta}$, $E_P[\Delta(Y, z_j, z_k)|Z^* = z_k] + (z_j - z_k)'\theta_0 \leq 0$ for all $z_j \neq z_k \in \mathcal{Z}$; (ii) $P(Z^* = z_k) > 0$ for all $1 \leq k \leq K$; (iii) $E_P[\Delta(Y, z_j, z_k)|Z^* = z_j] \neq E_P[\Delta(Y, z_j, z_k)|Z^* = z_k]$ for any $1 \leq j < k \leq K$; (iv) $\#\mathcal{A}(\theta, P) \leq d_\theta$ for all $\theta \in \Theta_0(P)$.*

Assumption F.3(i) guarantees $m(x,\theta)$ is bounded as required by Assumption 3.4(i). As in Assumption F.1(ii), $\Theta \subset \mathbf{R}^{d_\theta}$ is defined to be a sufficiently large sphere to ensure that $\Theta_0(P) \subset \Theta^o$, as demanded by Assumption 3.6(i). The gradient $\nabla F^{(i)}(\int m(x,\theta)dP(x)) \int \nabla_\theta m(x,\theta)dP(x)$ at each active constraint is of the form $(z_j - z_k)$ for some $z_j \neq z_k \in \mathcal{Z}$. Therefore, to ensure that Assumption 3.6(iv) holds, we must rule out that a $\theta \in \Theta_0(P)$ satisfies:

$$E[\Delta(Y, z_j, z_k)|Z^* = z_k] + (z_j - z_k)'\theta = 0 = E[\Delta(Y, z_k, z_j)|Z^* = z_j] + (z_k - z_j)'\theta , \tag{F.15}$$

which is guaranteed by Assumption F.4(iii). A consequence of Assumption F.4(iii) is that when Z^* has K points of support, it generates $K(K-1)$ constraints, of which at most $K(K-1)/2$ can be active. For this reason, Assumption F.3(iii) requires $K(K-1)/2 \geq d_\theta$, which together with Assumption F.3(iv) imply $\Theta_0(P)$ is bounded. Assumption F.4(i) is satisfied if the model is properly specified and implies $\Theta_0(P) \neq \emptyset$. Finally, Assumptions F.3(iv) and F.4(iv) together provide a sufficient condition for Assumption 3.6(iv) to be satisfied.

Remark F.1. The moment inequalities in (6) are a special case of a larger system implied by the optimality condition in (5). In particular, for any $\mathcal{F}$ measurable random variable V, equation (5) implies that for any $z_j \in \mathcal{Z}$:

$$E[((\psi(Y, z_j) - \psi(Y, Z^*)) + (z_j - Z^*)'\theta)g(V)] \leq 0 , \tag{F.16}$$

provided $g(V) \geq 0$ almost surely; see for example Ho (2009). Indeed, note (F.16) reduces to (6) by setting $V = Z^*$ and $g(V) = 1\{Z^* = z_k\}$. Unlike (6), however, it is not possible to write (F.16) as a linear inequality constraint with known slope for a general $g(V)$. On the other hand, (F.16) does satisfy Assumption 4.2. Therefore, Theorem 4.3 implies the "plug-in" estimator is still efficient for estimating $\nu_{|\mathbb{C}}(\cdot, \Theta_0(P))$ for any $\mathbb{C}$ satisfying Assumption 4.1. ■

Example 2.3 (Pricing Kernel)

For this example, we set $X = (Y, Z', U')'$ with $Y \in \mathbf{R}$, $Z \in \mathbf{R}^{dz}$ and $U \in \mathbf{R}^{dz}$, and hence $\mathcal{X} \subseteq \mathbf{R} \times \mathbf{R}^{dz} \times \mathbf{R}^{dz}$. Recall $\theta = (\rho, \gamma)' \in \mathbf{R}^2$, and to ensure the identified set is bounded, we impose the constraints $0 \leq \rho \leq \bar{\rho}$ and $0 \leq \gamma \leq \bar{\gamma}$ for some $\bar{\gamma} > 0$ and $\bar{\rho} > 0$. Formally, for a parameter space Θ, the identified set is given by:

$$\Theta_0(Q) \equiv \{\theta \in \Theta : \int (\frac{y^{-\gamma}z}{1+\rho} - u)dQ(x) \leq 0 \text{ and } \theta \in [0, \bar{\rho}] \times [0, \bar{\gamma}]\} . \tag{F.17}$$

In order to map this example into (2) and (3), we let A, $m_S : \mathcal{X} \times \Theta \to \mathbf{R}^{dz}$, and $F_S : \mathbf{R}^{dz} \to \mathbf{R}^{dz+4}$ be given by

$$m_S(x,\theta) = \frac{y^{-\gamma}z}{1+\rho} - u \qquad F_S(v) = (v', -\bar{\rho}, 0, -\bar{\gamma}, 0)' \qquad A' = \begin{bmatrix} 0'_{dz} & 1 & -1 & 0 & 0 \\ 0'_{dz} & 0 & 0 & 1 & -1 \end{bmatrix} , \tag{F.18}$$

where 0_{d_Z} stands for $0 \in \mathbf{R}^{d_Z}$. Given this notation, the constraints $1 \le i \le d_Z$ correspond to (7), while the restriction $\theta \in [0,\bar{\rho}] \times [0,\bar{\gamma}]$ is imposed in the constraints $d_Z + 1 \le i \le d_Z + 4$. Therefore, we obtain the representation

$$\Theta_0(Q) = \{\theta \in \Theta : A\theta + F_S(\int m_S(x,\theta)dQ(x)) \le 0\} . \tag{F.19}$$

The following conditions are sufficient for verifying Assumptions 3.2-3.6 in Example 2.3.

Assumption F.5. *(i) $\mathcal{X} \subseteq [\epsilon_0,\infty)\times \mathbf{R}_+^{d_Z} \times \mathbf{R}^{d_Z}$ for some $\epsilon_0 > 0$; (ii) $\mathcal{X}$ is compact; (iii) $\Theta \equiv [-1/2, 2\bar{\rho}]\times[-1/2,2\bar{\gamma}]$.*

Assumption F.6. *(i) $E[\frac{Y^{-\gamma}Z}{1+\rho} - U] \le 0$ for some $\theta \in [0,\bar{\rho}] \times [0,\bar{\gamma}]$; (ii) $P(Z^{(i)} > 0) > 0$ for all $1 \le i \le d_Z$; (iii) For all $(\rho,\gamma)' = \theta \in \Theta_0(P)$, and $\{i,j\} \subseteq \mathcal{A}(\theta,P)$ with $1 \le i < j \le d_Z+2$, $E[Y^{-\gamma}(Z^{(i)} - \pi_{i,j}Z^{(j)})\log(Y)] \neq 0$, where $\pi_{i,j} = E[U^{(i)}]/E[U^{(j)}]$ if $j \le d_Z$ and $\pi_{i,j} = 0$ otherwise; (iv) $\#\mathcal{A}(\theta,P) \le 2$ for all $\theta \in \Theta_0(P)$.*

Assumption F.5(i) requires Y, the ratio of future over current consumption, to be bounded away from zero. Together with compactness of $\mathcal{X}\times\Theta$, Assumption F.5(i) ensures $m : \mathcal{X}\times\Theta \to \mathbf{R}^{2d_Z+4}$ is bounded and differentiable, as required by Assumption 3.4. The constraint $\theta \in [0,\bar{\rho}] \times [0,\bar{\gamma}]$ can be interpreted as imposing restrictions defining the parameter space of interest (see Remark 3.4). However, our arguments require regularity of m in a neighborhood of $\Theta_0(P)$, and for this reason Assumption 3.6(i) further demands we may define a set Θ such that $\Theta_0(P) \subset \Theta^o$. In this example, this is easily accomplished through Assumption F.5(iii) – alternatively, for example, we could have set $\Theta = [-\delta, \bar{\rho}+\delta] \times [-\delta, \bar{\gamma}+\delta]$ for any $0 < \delta < 1$. Assumption F.6(i) implies $\Theta_0(P) \neq 0$, and is satisfied if the model is properly specified. In turn, Assumption F.6(ii) is necessary for $\theta \mapsto F_S^{(i)}(\int m_S(x,\theta)dP(x))$ to be strictly convex for $1 \le i \le d_Z$. Finally, Assumptions F.6(iii)-(iv) are equivalent to Assumption 3.6(iv) in this model. Unfortunately, unlike in the linear models of Examples 2.1 and 2.2, the gradients of constraints $1 \le i \le d_Z$ depend on P, and as a result the requirement on P is more complex.

Proposition F.2. *In Example 2.3, Assumptions F.5 and F.6 imply Assumptions 3.2-3.6.*

Proof: Assumption 3.2(i) is implied by Assumption F.5(iii), while Assumption 3.3(i) has already been verified in (F.18) and (F.19). Moreover, since $y \ge \epsilon_0 > 0$ for all $x \in \mathcal{X}$ and $\rho \ge -1/2$ for all $(\rho,\gamma) = \theta \in \Theta$, and $\mathcal{X}\times\Theta$ is compact by Assumption F.5(i)-(ii), it also follows that $m_S(x,\theta)$ is uniformly bounded on $(x,\theta) \in \mathcal{X}\times\Theta$. Therefore, $m(x,\theta) = (m_S(x,\theta)', \theta'A')'$ implies Assumption 3.4(i) also holds. Next, note by direct calculation that:

$$\nabla_\theta m_S(x,\theta) = \begin{bmatrix} -\frac{y^{-\gamma}z}{(1+\rho)^2} & -\frac{y^{-\gamma}\log(y)z}{(1+\rho)} \end{bmatrix} , \tag{F.20}$$

and hence since $\rho \ge -1/2$ and $y \ge \epsilon_0$ by Assumptions F.5(i) and F.5(iii), it follows that $(x,\theta) \mapsto \nabla_\theta m_S(x,\theta)$ is uniformly bounded in $\mathcal{X}\times\Theta$. Assumption 3.4(ii) then follows from $\nabla_\theta m(x,\theta) = (\nabla_\theta m_S(x,\theta)', A')'$. Moreover, (F.20) further implies $(\theta,x) \mapsto \nabla_\theta m(x,\theta)$ is continuous on $\mathcal{X}\times\Theta$. However, by compactness of $\mathcal{X}\times\Theta$, $(\theta,x) \mapsto \nabla_\theta m(x,\theta)$ is uniformly continuous, and therefore $\theta \mapsto \nabla_\theta m(x,\theta)$ is equicontinuous in $x \in \mathcal{X}$, verifying Assumption 3.4(iii). Finally, employing $m(x,\theta) = (m_S(x,\theta)', \theta'A)'$, and $F(\int m(x,\theta)dQ(x)) = A\theta + F_S(\int m_S(x,\theta)dQ(x))$ we obtain:

$$\nabla F(v) = \begin{bmatrix} I_{d_Z} & \vdots\, I_{d_Z+4} \\ 0_{4,d_Z} & \end{bmatrix} , \tag{F.21}$$

where I_k denotes the $k\times k$ identity matrix, and $0_{4,d_Z}$ is a $4 \times d_Z$ matrix of zeroes. From (F.21) it follows that Assumptions 3.5(i)-(ii) hold with $V_0 = \mathbf{R}^{2d_Z+4}$.

To verify Assumption 3.6, first observe Assumption F.2(i) directly imposes $\Theta_0(P) \neq \emptyset$. Moreover, since $\Theta_0(P) \subseteq [0,\bar{\rho}]\times[0,\bar{\gamma}] \subset (-1/2, 2\bar{\rho}) \times (-1/2, 2\bar{\gamma}) = \Theta^o$ by Assumption F.5(iii), it follows that Assumption 3.6(i) holds. To verify Assumption 3.6(ii), first note that by (F.18) $\mathcal{S}_i = \{1,2\}$ for $1 \le i \le d_Z$ and $\mathcal{S}_i = \emptyset$ for $d_Z+1 \le i \le d_Z+4$. Thus, we need only show $\theta \mapsto \int m_S^{(i)}(x,\theta)dQ(x)$ is strictly convex for all $1 \le i \le d_Z$ and Q in a suitable neighborhood of P. To this end, first exploit that $\rho \ge -1/2$ for all $(\rho,\gamma) \in \Theta$ and $y \ge \epsilon_0$ for all $x \in \mathcal{X}$ to deduce that

$$\nabla_\theta^2 m_S^{(i)}(x,\theta) = \begin{bmatrix} \frac{2y^{-\gamma}z^{(i)}}{(1+\rho)^3} & \frac{y^{-\gamma}\log(y)z^{(i)}}{(1+\rho)^2} \\ \frac{y^{-\gamma}\log(y)z^{(i)}}{(1+\rho)^2} & \frac{y^{-\gamma}\log^2(y)z^{(i)}}{(1+\rho)} \end{bmatrix} , \tag{F.22}$$

for any $(x,\theta) \in \mathcal{X} \times \Theta$ and $1 \leq i \leq d_Z$. By (F.22), $\nabla^2_\theta m_S^{(i)}(x,\theta)$ is positive definite for any $x \in \mathcal{X}$ such that $z^{(i)} > 0$. Hence, since $z^{(i)} \geq 0$ on $\mathcal{X}$, and $m_S^{(i)}(x,\theta) = -u$ whenever $z^{(i)} = 0$, we conclude that for any $\lambda \in (0,1)$ and $1 \leq i \leq d_Z$

$$\int m_S^{(i)}(x, \lambda\theta_1 + (1-\lambda)\theta_2) dQ(x) < \lambda \int m_S^{(i)}(x,\theta_1) dQ(x) + (1-\lambda) \int m_S^{(i)}(x,\theta_2) dQ(x) \tag{F.23}$$

provided that $Q \in \mathbf{M}$ satisfies $Q(Z^{(i)} > 0) > 0$. However, by Assumption F.6(ii) $P(Z^{(i)} > 0) > 0$ for all $1 \leq i \leq d_Z$. Hence, for each $1 \leq i \leq d_Z$ there exists a neighborhood $N_i(P) \subseteq \mathbf{M}$ in the τ-topology such that $Q(Z^{(i)} > 0) > 0$ for all $Q \in N_i(P)$. Therefore, by (F.23), Assumption 3.6(ii) then holds with $N(P) = \bigcap_i N_i(P)$. In turn, Assumption 3.6(iii) trivially holds since $V_0 = \mathbf{R}^{2d_Z+4}$. Finally, to verify Assumption 3.6(iv) first note

$$\nabla F\left(\int m(x,\theta)dP(x)\right) \int \nabla_\theta m(x,\theta) dP(x) = \begin{bmatrix} -\int \frac{y^{-\gamma} z'}{(1+\rho)^2} dP(x) & 1 & -1 & 0 & 0 \\ -\int \frac{y^{-\gamma}\log(y) z'}{(1+\rho)} dP(x) & 0 & 0 & 1 & -1 \end{bmatrix}' \tag{F.24}$$

by direct calculation and (F.21). Since $P(Z^{(i)} > 0) > 0$ for all $1 \leq i \leq d_Z$ and $y \geq \epsilon_0 > 0$ for all $x \in \mathcal{X}$, we must have $E[Y^{-\gamma} Z^{(i)}] > 0$. Therefore, $\nabla F^{(i)}(\int m(x,\theta)dP(x)) \int \nabla_\theta m(x,\theta) dP(x) \neq 0$ for all $1 \leq i \leq d_Z$, and thus $\{\nabla F^{(i)}(\int m(x,\theta)dP(x)) \int \nabla_\theta m(x,\theta) dP(x)\}_{i \in \mathcal{A}(\theta,P)}$ are linearly independent if $\mathcal{A}(\theta,P)$ is either empty or singleton valued. Hence, by Assumption F.6(iv), we need only consider the case $\mathcal{A}(\theta,P) = \{i,j\}$ with $i \neq j$. However, note that if $j \in \{d_Z+3, d_Z+4\}$, then $i \leq d_Z+2$ (since the d_Z+3 and d_Z+4 constraint cannot simultaneously bind), and by (F.24) and $E[Y^{-\gamma} Z^{(k)}] > 0$ Assumption 3.6(iv) is satisfied. Finally, for the case $1 \leq i < j \leq d_Z+2$ Assumption 3.6(iv) follows by direct calculation, Assumption F.6(iii) and exploiting that if $i \in \mathcal{A}(\theta,P)$ and $i \leq d_Z$, then $E[Y^{-\gamma} Z^{(i)}] = (1+\rho) E[U^{(i)}]$. ■

Example 2.4 (Participation Constraint)

In order to write this example in the form of (2) and (3), let $X = (C, W, L, Z')'$ with $(C,W,L) \in \mathbf{R}^3_+$ and $Z \in \mathbf{R}^{d_Z}_+$. We denote the parameter $\theta = (\alpha,\beta)' \in \mathbf{R}^2$ and we ensure $\Theta_0(P)$ is bounded by imposing the constraints $0 \leq \alpha \leq \bar{\alpha}$ and $0 \leq \beta \leq \bar{\beta}$ with $\bar{\alpha} > 0$ and $\bar{\beta} > 0$. For a parameter space Θ, then define the identified set

$$\Theta_0(Q) \equiv \{\theta \in \Theta : \int \left(\frac{w}{c-\alpha} - \frac{\beta}{l}\right) z dQ(x) \leq 0 \text{ and } \theta \in [0,\bar{\alpha}] \times [0,\bar{\beta}]\} . \tag{F.25}$$

Further let $m_S(x,\theta) = (z'w/(c-\alpha), z'/l)'$ and define a $(d_Z+4) \times 2$ matrix A and $F_S : \mathbf{R}^{2d_Z} \to \mathbf{R}^{d_Z+4}$ by:

$$F_S^{(i)}(v) = \begin{cases} \frac{v^{(i)}}{v^{(d_Z+i)}} & \text{if } 1 \leq i \leq d_Z \\ -\bar{\alpha} & \text{if } i = d_Z+1 \\ 0 & \text{if } i \in \{d_Z+2, d_Z+4\} \\ -\bar{\beta} & \text{if } i = d_Z+3 \end{cases} \qquad A' = \begin{bmatrix} 0'_{d_Z} & 1 & -1 & 0 & 0 \\ -1'_{d_Z} & 0 & 0 & 1 & -1 \end{bmatrix} \tag{F.26}$$

where 1_{d_Z} is a vector of ones in $\mathbf{R}^{d_Z}$ and recall 0_{d_Z} denotes $0 \in \mathbf{R}^{d_Z}$. Thus, for $1 \leq i \leq d_Z$ we obtain the constraint

$$F^{(i)}\left(\int m(x,\theta)dP(x)\right) = -\beta + \frac{E[WZ^{(i)}/(C-\alpha)]}{E[Z^{(i)}/L]} , \tag{F.27}$$

while constraints $d_Z+1 \leq i \leq d_Z+4$ impose $\theta \in [0,\bar{\alpha}] \times [0,\bar{\beta}]$. Given this notation, we may then rewrite:

$$\Theta_0(Q) = \{\theta \in \Theta : A\theta + F_S\left(\int m_S(x,\theta)dQ(x)\right) \leq 0\} . \tag{F.28}$$

Assumptions F.7 and F.8 impose sufficient conditions for verifying Assumptions 3.2-3.6.

Assumption F.7. *(i) $\mathcal{X} \subseteq [\epsilon_0,\infty) \times \mathbf{R}_+ \times [\epsilon_1,+\infty) \times \mathbf{R}^{d_Z}_+$ for some $\epsilon_0 > \bar{\alpha}$ and $\epsilon_1 > 0$; (ii) $\mathcal{X}$ is compact; (iii) $\Theta \equiv [-\delta_0, \bar{\alpha}+\delta_0] \times [-\delta_0, \bar{\beta}+\delta_0]$ for some $0 < \delta_0 < (\epsilon_0 - \bar{\alpha})$.*

Assumption F.8. *(i) $E[(\frac{W}{C-\alpha} - \frac{\beta}{L})Z] \leq 0$ for some $\theta \in [0,\bar{\alpha}] \times [0,\bar{\beta}]$; (ii) $P(WZ^{(i)} > 0) > 0$ for all $1 \leq i \leq d_Z$; (iii) For all $(\alpha,\beta)' = \theta \in \Theta_0(P)$, and $\{i,j\} \subseteq \mathcal{A}(\theta,P)$ with $1 \leq i < j \leq d_Z$, $E[\frac{W}{(C-\alpha)^2}(Z^{(i)} - \pi_{i,j} Z^{(j)})] \neq 0$, where $\pi_{i,j} = E[\frac{Z^{(i)}}{L}]/E[\frac{Z^{(j)}}{L}]$; (iv) $\#\mathcal{A}(\theta,P) \leq 2$ for all $\theta \in \Theta_0(P)$.*

In Assumptions F.7(i) and F.7(iii) we impose that $C - \alpha$ and L be bounded away from zero, as required for $m(x, \theta)$ to be bounded and utility to remain finite (recall $u(C, L) = \log(C - \alpha) + \beta \log(L)$). As in Example 2.3, in Assumption F.7(iii) we define Θ to be an expansion of the parameter constraints $\theta \in [0, \bar{\alpha}] \times [0, \bar{\beta}]$. Assumption F.8(i) ensures $\Theta_0(P) \neq \emptyset$, while Assumption F.8(ii) is required so that constraints $1 \leq i \leq d_Z$ are strictly convex in α. Finally, Assumptions F.8(iii)-(iv) are necessary and sufficient for P to satisfy Assumption 3.6(iv) in this model. As in Example 2.3, the gradients of constraints $1 \leq i \leq d_Z$ depend on P, which leads to a more complex requirement than was necessary in Examples 2.1 and 2.2.

Proposition F.3. *In Example 2.4, Assumptions F.7 and F.8 imply Assumptions 3.2-3.6.*

Proof: Assumption 3.2(i) is implied by Assumption F.7(iii), while Assumption 3.3(i) was already verified in (F.28). Moreover, compactness of $\mathcal{X} \times \Theta$ implies wz and z are uniformly bounded, while $c \geq \epsilon_0 > \bar{\alpha} + \delta_0 \geq \alpha$ and $l \geq \epsilon_1$ implies $(c - \alpha)^{-1}$ and l^{-1} are uniformly bounded as well. Therefore, $m_S(x, \theta) = (z'w/(c - \alpha), z'/l)'$ is uniformly bounded in $(x, \theta) \in \mathcal{X} \times \Theta$ and hence so is $m(x, \theta) = (m_S(x, \theta)', \theta' A')'$, which verifies Assumption 3.4(i). Similarly,

$$\nabla_\theta m_S(x, \theta) = \begin{bmatrix} \frac{wz}{(c-\alpha)^2} & 0_{d_z} \\ 0_{d_Z} & 0_{d_Z} \end{bmatrix} \tag{F.29}$$

is also bounded, which together with $\nabla_\theta m(x, \theta) = (\nabla_\theta m_S(x, \theta)', A')'$ implies Assumption 3.4(ii) holds as well. In turn, by compactness of $\mathcal{X} \times \Theta$ and (F.29), $(x, \theta) \mapsto \nabla_\theta m(x, \theta)$ is uniformly continuous on $\mathcal{X} \times \Theta$ and therefore $\theta \mapsto \nabla_\theta m(x, \theta)$ is equicontinuous in $x \in \mathcal{X}$ as demanded by Assumption 3.4(iii). Next, let $\eta_0 < \inf_k E[Z^{(k)}/L]$ and note we may set $\eta_0 > 0$ due to Assumption F.8(ii) and $P(W \geq 0) = 1$ by definition of $\mathcal{X}$. Similarly, let $\sup_{\mathcal{X} \times \Theta} \|m(x, \theta)\| < M_0$, and note that since Assumption 3.4(ii) holds, we may set $M_0 < \infty$. Then defining

$$V_0 \equiv (-M_0, M_0)^{d_Z} \times (\eta_0, M_0)^{d_Z} \times (-M_0, M_0)^{d_Z + 4} , \tag{F.30}$$

and noting that $F(v)$ is differentiable unless $v^{(i)} = 0$ for some $d_Z + 1 \leq i \leq 2d_Z$, it follows that Assumption 3.5(i) holds. In addition, since ∇F is continuous on the closure of V_0 and such closure is compact, it follows that Assumption 3.5(ii) holds as well.

In order to verify P satisfies Assumption 3.6, first note that by Assumptions F.7(iii) and F.8(i), $\emptyset \neq \Theta_0(P) \subseteq [0, \bar{\alpha}] \times [0, \bar{\beta}] \subset (-\delta_0, \bar{\alpha} + \delta_0) \times (-\delta_0, \bar{\beta} + \delta_0) = \Theta^o$ which verifies Assumption 3.6(i). Next observe $\mathcal{S}_i = \emptyset$ for $d_Z + 1 \leq i \leq d_Z + 4$, and $\mathcal{S}_i = \{1\}$ for $1 \leq i \leq d_Z$. Therefore, to show Assumption 3.6(ii) holds it suffices to establish

$$F_S^{(i)}(\int m(x, \theta) dQ(x)) = \frac{\int (wz^{(i)}/(c - \alpha)) dQ(x)}{\int (z^{(i)}/l) dQ(x)} , \tag{F.31}$$

is strictly convex in α for all Q in an appropriate neighborhood of P. However, by Assumptions F.7(i) and F.8(ii) $E[Z^{(i)}/L] > 0$ and $E[WZ^{(i)}] > 0$, and therefore there exists a neighborhood $N_i(P) \subseteq \mathbf{M}$ such that $\int (z^{(i)}/l) dQ(x) > 0$ and $\int wz^{(i)} dQ(x) > 0$ for all $Q \in N_i(P)$. Letting $N(P) = \bigcap_i N(P)$ and noting that $Q(C - \alpha > 0) = 1$ for all $\alpha \in [0, \bar{\alpha}]$ and $Q \in \mathbf{M}$ by Assumption F.7(i), we obtain that $\alpha \mapsto F_S^{(i)}(\int m(x, \theta) dQ(x))$ is indeed strictly convex for all $Q \in N(P)$, thus verifying Assumption 3.6(ii). In turn, Assumption 3.6(iii) is also satisfied by construction of V_0 in (F.30) and definitions of η_0 and M_0. Finally, note that by (F.29) and direct calculation

$$\nabla F(\int m(x, \theta) dP(x)) \int \nabla_\theta m(x, \theta) dP(x) = \begin{bmatrix} \frac{E[WZ'/(C-\alpha)^2]}{E[Z'/L]} & 1 & -1 & 0 & 0 \\ -1'_{d_Z} & 0 & 0 & 1 & -1 \end{bmatrix}' . \tag{F.32}$$

Hence, since $E[WZ^{(i)}/(C - \alpha)] > 0$ and $E[Z^{(i)}/L] > 0$ for all $\alpha \in [0, \bar{\alpha}]$ and $1 \leq i \leq d_Z$ by Assumptions F.7(i) and F.8(ii), (F.32) implies $\nabla F^{(i)}(\int m(x, \theta) dP(x)) \int \frac{\partial}{\partial \theta^{(j)}} m(x, \theta) dP(x) \neq 0$ for any $1 \leq i \leq d_Z$ and $j \in \{1, 2\}$. As a result, it follows that $\{\nabla F^{(i)}(\int m(x, \theta) dP(x)) \int \nabla_\theta m(x, \theta) dP(x)\}_{i \in \mathcal{A}(\theta, P)}$ are linearly independent whenever $\mathcal{A}(\theta, P)$ is empty or a singleton, and also when $\{i, j\} = \mathcal{A}(\theta, P)$ with $i < j$ and $j \geq d_Z + 1$. Therefore, by Assumption F.8(iv), to verify Assumption 3.6(iv) it only remains to consider the case $\{i, j\} = \mathcal{A}(\theta, P)$ with $j \leq d_Z$. However, in this instance $\{\nabla F^{(i)}(\int m(x, \theta) dP(x)) \int \nabla_\theta m(x, \theta) dP(x)\}_{i \in \mathcal{A}(\theta, P)}$ are linearly independent by result (F.32), Assumption F.8(iv) and direct calculation, and hence P satisfies Assumption 3.6(iv) as well. ■

Appendix G - Simulation Evidence

In this Appendix, 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. For comparison purposes, we also include the results of employing the uniformly valid procedures proposed in Andrews and Soares (2010) and Bugni (2010).

For our design, we let $Z_i \equiv (Z_i^{(1)}, Z_i^{(2)})'$ where $Z_i^{(1)} = 1$ is a constant and $Z_i^{(2)}$ 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 according to:

$$Y_i = Z_i'\theta_0 + \epsilon_i \qquad i = 1, \ldots, n \; , \tag{G.1}$$

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

$$\begin{aligned} Y_{L,i} &= Y_i - C - V_i(Z_i^{(2)})^2 \qquad i = 1, \ldots, n \\ Y_{U,i} &= Y_i + C + V_i(Z_i^{(2)})^2 \qquad i = 1, \ldots, n \; , \end{aligned} \tag{G.2}$$

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, $\Theta_0(P)$ 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 (see also (F.2)). 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, while C controls the diameter of the identified set with point identification occurring at $C = 0$ – see Figure 1. Throughout our simulation study we will examine specifications with $C \in \{0.1, 0.5, 1\}$ and $K \in \{5, 9, 15\}$, with the latter corresponding to 10, 18 and 30 moment inequalities respectively. Heuristically, high values of K or low values of C yield specifications where P is closer to violating Assumption 3.6(iv). In such instances, we therefore expect our asymptotic results to provide a less reliable approximation to finite sample distributions, while uniform procedures should remain accurate.

We first compare the performance of the efficient set estimator $\hat{\Theta}_n = co\{\Theta_0(\hat{P}_n)\}$ (see (19)) with that of:

$$\hat{\Theta}_n(\tau_n) \equiv \{\theta \in \Theta : F^{(i)}(\int m(x, \theta) d\hat{P}_n(x)) \leq \frac{\tau_n}{\sqrt{n}} \hat{\sigma}_n^{(i)} \text{ for } i = 1, \ldots, d_F\} \; , \tag{G.3}$$

where $(\hat{\sigma}_n^{(i)})^2$ is a consistent estimator for the asymptotic variance of constraint number i.[15] Chernozhukov et al. (2007) and Bugni (2010) show $\hat{\Theta}_n(\tau_n)$ is a consistent estimator for $\Theta_0(P)$ under the Hausdorff metric provided that $\tau_n/\sqrt{n} \downarrow 0$. Notice in particular that the efficient estimator $\hat{\Theta}_n$ corresponds to setting $\tau_n = 0$, and is therefore by construction always smaller than $\hat{\Theta}_n(\tau_n)$ whenever $\tau_n > 0$. This is not necessarily a favorable property, however, since an estimator that is too small may perform poorly in terms of Hausdorff distance to $\Theta_0(P)$. For example, in certain specifications we find in many replications that $\hat{\Theta}_n(\tau_n) = \emptyset$ for values of $\tau_n \in \{0, \log(\log(n))\}$, in which case the Hausdorff distance to $\Theta_0(P)$ is set to equal infinity. Table 1 reports the proportion of replications for which this event occurs in each specification. As expected, the most problematic specifications are those with many moment inequalities ($K = 15$) and $\Theta_0(P)$ near point identification ($C = 0.1$).

Table 2 reports the median of the Hausdorff distance between the different set estimators and $\Theta_0(P)$ across replications – see Remark G.1 for computational details. We report median, rather than mean, Hausdorff distance because $d_H(\hat{\Theta}_n(\tau_n), \Theta_0(P))$ is infinite in replications for which $\hat{\Theta}_n(\tau_n) = \emptyset$. As expected, the median Hausdorff distance decreases with sample size across all specifications and choices of τ_n. Interestingly, for $\tau_n \in \{\log(\log(n)), \log(n), n^{1/8}, n^{1/4}\}$ the performance of $\hat{\Theta}_n(\tau_n)$ is completely insensitive to the choice of C across all specifications, while the performance of the efficient estimator is only sensitive to the value of C when many moment inequalities are present ($K = 15$). In contrast, the median Hausdorff distance of all estimators deteriorates as the

[15] In particular, for $\bar{m}_n(\theta) \equiv \int m(x, \theta) d\hat{P}_n(x)$, and $\hat{\Omega}_n(\theta) \equiv \int (m(x, \theta) - \bar{m}_n(\theta))(m(x, \theta) - \bar{m}_n(\theta))' d\hat{P}_n(x)$, we let $(\hat{\sigma}_n^{(i)})^2 \equiv \nabla F^{(i)}(\int m(x, \theta) d\hat{P}_n(x)) \hat{\Omega}_n(\theta) \nabla F^{(i)}(\int m(x, \theta) d\hat{P}_n(x))'$. It is easy to verify $\hat{\sigma}_n^{(i)}$ does not depend on θ.

number of moment inequalities increases. Remarkably, across almost all specifications the median Hausdorff distance is monotonically increasing in τ_n, with the efficient estimator outperforming all the alternative estimators.[16] The notable exception is the specification $K = 15$, $C = 0.1$ and $n = 200$, in which the median Hausdorff distance of the efficient estimator is infinite due to $\Theta_0(\hat{P}_n)$ being empty in over half the replications (see Table 1).

Next, we examine the performance of inferential procedures based on the semiparametric efficient estimator and compare it to that of alternative methods that are asymptotically valid uniformly in P. To this end, we first consider the construction of confidence regions $\mathcal{C}_n$ for the identified set $\Theta_0(P)$ satisfying the coverage requirement:

$$\liminf_{n\to\infty} P(\Theta_0(P) \subseteq \mathcal{C}_n) \geq 1 - \alpha \ . \tag{G.4}$$

Following the discussion in Example 5.1, we employ the efficient estimator to obtain a confidence region satisfying (G.4) by using a construction proposed in Beresteanu and Molinari (2008) – see Remark G.2 for computational details. Additionally, we also obtain confidence regions satisfying (G.4) by utilizing a criterion function based approach, as developed in Chernozhukov et al. (2007) and Bugni (2010). Specifically, defining the criterion function:

$$Q_n(\theta) \equiv \max_{1\leq i\leq d_F} \frac{1}{\hat{\sigma}_n^{(i)}}(F^{(i)}(\int m(x,\theta)d\hat{P}_n(x)))_+ \ , \tag{G.5}$$

we examine confidence regions of the form $CS_n(\tau_n) \equiv \{\theta \in \Theta : Q_n(\theta) \leq \hat{c}^B_{1-\alpha}(\tau_n)/\sqrt{n}\}$, where $\hat{c}^B_{1-\alpha}(\tau_n)$ is the critical value proposed in Bugni (2010) – see Remark G.3. Employing the maximum, rather than the sum, across constraints in defining Q_n implies $CS_n(\tau_n)$ is a convex polygon, which greatly simplifies our computations. All bootstrap procedures employed 200 replications in computing critical values.

Table 3 reports the coverage probabilities of the different confidence regions under alternative values of (n, K, C) for a nominal coverage of 0.95. A confidence region based on the efficient estimator is considered to have failed to cover $\Theta_0(P)$ in any replication for which $\Theta_0(\hat{P}_n) = \emptyset$. Similarly, the criterion based confidence region is considered to have failed to cover $\Theta_0(P)$ whenever $\hat{\Theta}_n(\tau_n) = \emptyset$ – see Remark G.3. As in Table 2, the performance of the confidence region based on the efficient estimator is more sensitive to K than to C. In specifications with 10 moment inequalities ($K = 5$), the actual coverage is always close to its nominal level, while under 30 moment inequalities ($K = 15$) size distortions upwards of 5% remain even for $n = 1000$. Unsurprisingly, the most severe undercoverage occurs in specifications for which $\Theta_0(\hat{P}_n) = \emptyset$ in a large number of replications ($K = 15, C = 0.1$). In contrast, the criterion based confidence regions have actual coverage above the nominal level for all specifications. The coverage probability is closest to the nominal level under 10 moment inequalities ($K = 5$), but can be quite conservative for larger values of the slackness parameter τ_n ($\tau_n \in \{\log(n), n^{1/4}\}$).

In Table 4 we report the median Hausdorff distance between the different confidence regions and the identified set $\Theta_0(P)$. For specifications in which all confidence regions control size, the median Hausdorff distance of the confidence region based on the efficient estimator is always smaller than that of its competitors. These results suggest that while the criterion based confidence regions can deliver uniform size control, they can also underperform when our asymptotic results provide an accurate approximation to finite sample distributions. Finally, in Table 9 we tabulate the median computation time in seconds for each confidence region. The computational time of all approaches is small, but longest for the confidence region based on the efficient estimator. It is worth noting that the Lagrange multipliers $\lambda(p, \hat{P}_n)$ and maximizers $\hat{\theta}(p)$ needed to construct $G^*_n(p)$ (as in (22)) are by-products of computing $\nu(p, \Theta_0(\hat{P}_n))$. As a result, simulating the distribution of G^*_n only requires sampling $\{W_i\}_{i=1}^n$, which significantly reduces computation time relative to a procedure that recomputes the support function in each bootstrap iteration.

[16] Note that for all the values of n we consider $\log(\log(n)) < n^{\frac{1}{8}} < \log(n) < n^{\frac{1}{4}}$.

We further evaluate the size and power of the test based on $J_n(\theta)$ (see (27)) for the null hypothesis:

$$H_0 : \theta \in \Theta_0(P) \qquad\qquad H_1 : \theta \notin \Theta_0(P) \ . \tag{G.6}$$

In order to make size control nontrivial, we let θ be a boundary point of $\Theta_0(P)$. In particular, for the vectors:

$$\theta_F \equiv (\nu((1,0),\Theta_0(P)),0) \qquad\qquad \theta_K \equiv (0,\nu((0,1),\Theta_0(P))) \tag{G.7}$$

we consider the hypothesis testing problem in (G.6) when $\theta \in \{\theta_F, \theta_K\}$. Notice, that θ_F and θ_K are respectively points in a "flat face" and at a "kink" of $\Theta_0(P)$ for all values of (C, K) (see Figure 1). Thus, θ_F is supported by a unique hyperplane while θ_K is supported by multiple hyperplanes, which implies Theorem 5.3 applies to the former but not the latter. For comparison purposes we also examine the performance of the generalized moment selection procedure developed in Andrews and Soares (2010). Specifically, for $\theta \in \{\theta_F, \theta_K\}$ we consider a test that rejects the null hypothesis in (G.6) whenever $\sqrt{n}Q_n(\theta) > \hat{c}^{AS}_{1-\alpha}(\theta)$ for a bootstrap critical value $\hat{c}^{AS}_{1-\alpha}(\theta)$ – see Remark G.4. Both procedures require a choice of slackness parameter (see (29)), which we pick select the set $\{\log(\log(n)), \log(n), n^{1/8}, n^{1/4}\}$.

Tables 5 and 6 report the actual size of tests of (G.6) for a nominal size of 0.05 and $\theta \in \{\theta_F, \theta_K\}$. For tests based on the efficient estimator, we considered the null hypothesis in (G.6) to be rejected in any replication for which $\Theta_0(\hat{P}_n) = \emptyset$. The performance of the tests for (G.6) when $\theta = \theta_F$ are similar to those of the confidence regions for $\Theta_0(P)$ (Table 3). In particular, the test based on the efficient estimator provides accurate size control under ten moment inequalities ($K = 10$), but can fail to do so under 30 moment inequalities ($K = 15$). With the exception of those specifications in which $\Theta_0(\hat{P}_n) = \emptyset$ in a significant number of replications, however, the size distortions are not as severe as those in Table 3. In contrast, the test of Andrews and Soares (2010) always provides adequate size control, though it can sometimes be severely conservative, for instance for $K = 15$ and $C = 0.1$. The patterns when $\theta = \theta_K$ are similar, though all tests have a weakly lower rejection rate than when $\theta = \theta_F$ in a majority of the specifications. As a result, for larger values of κ_n ($\kappa_n \in \{\log(n), n^{1/4}\}$), the test based on the efficient estimator delivers adequate size control in all specification except those for which $\Theta_0(\hat{P}_n) = \emptyset$ in a large proportion of replications (see Table 1).

In order to evaluate the local power of the proposed tests, we further test (G.6) when θ is of the form

$$\theta = \theta_C + \frac{h}{\sqrt{n}}\theta_A \tag{G.8}$$

where $\theta_C \in \{\theta_F, \theta_K\}$, and $\theta_A = (1,0)$ if $\theta_C = \theta_F$ and $\theta_C = (0,1)$ otherwise. It can be verified by direct calculation that $h/\sqrt{n} = \inf_{\tilde{\theta}\in\Theta_0(P)} \|\theta - \tilde{\theta}\|$ whenever $h \geq 0$, and hence h controls the distance of the local alternative to the identified set. Tables 7 and 8 report rejection probabilities for tests with a nominal size of 0.05. We focus on specifications with $K \in \{5, 9\}$ so that both tests provide adequate size control, and ignore specifications with $n = 500$ for conciseness. Notice that results with $h = 0$ correspond to the actual size of the test. For local deviations away from $\theta = \theta_F$ (Table 7), the test based on the efficient estimator is more powerful than its competitors in almost all specifications, and the pattern is robust to the choice of slackness parameters. Interestingly, all tests are more powerful in detecting local deviations away from $\theta = \theta_K$ (Table 8) than from $\theta = \theta_F$. However, in this instance the tests are also more sensitive to the choice of slackness parameters κ_n and τ_n. As a result, the power comparison of tests in detection deviations away from $\theta = \theta_K$ is not as conclusive as in Table 7.

In the results reported in Tables 2-6 the performance of statistics based on the efficient estimator is always worst in specifications for which $\Theta_0(\hat{P}_n) = \emptyset$ in a large number of replications. However, upon finding $\Theta_0(\hat{P}_n) = \emptyset$ it is evident that our asymptotic approximation is inadequate – in fact, the developed statistics cannot even be computed. For completeness, it is therefore also important to examine the performance of these procedures conditional on having found $\Theta_0(\hat{P}_n) \neq \emptyset$. These results are reported in Table 10. Surprisingly, the procedures perform well, with our confidence intervals and tests actually being conservative in such instances. We emphasize, however, that there is no reason to expect the results of Table 10 to hold in generality. Thus, special care should be taken in applying procedures based on the efficient estimator whenever there is reason to doubt the relevance of Assumption 3.6(iv).

Remark G.1. Since each function $\theta \mapsto F^{(i)}(\int m(x,\theta)d\hat{P}_n(x))$ is linear for all $1 \leq i \leq d_F$, the sets $\hat{\Theta}_n(\tau_n)$ are convex polygons. Moreover, their support functions are easily computable through the optimization problem:[17]

$$\nu(p, \hat{\Theta}_n(\tau_n)) = \sup_{\theta}\{\langle p, \theta\rangle \text{ s.t. } F^{(i)}(\int m(x,\theta)d\hat{P}_n(x)) \leq \frac{\tau_n}{\sqrt{n}}\hat{\sigma}^{(i)} \text{ for } i = 1, \ldots d_F\} \ . \tag{G.9}$$

In our simulations, we approximate $\mathbb{S}^2$ by letting $\mathcal{G}$ be a 100 point grid of $[-\pi, \pi]$, and considering the vectors:

$$p(\gamma) \equiv (\sin(\gamma), \cos(\gamma)) \tag{G.10}$$

for $\gamma \in \mathcal{G}$. Exploiting (9), we then approximate $d_H(\hat{\Theta}_n(\tau_n), \Theta_0(P))$ by $\max_{\gamma\in\mathcal{G}} |\nu(p(\gamma), \hat{\Theta}_n(\tau_n)) - \nu(p(\gamma), \Theta_0(P))|$. ■

Remark G.2. Because in this context all constraints are linear in θ, the support function has the dual representation:

$$\nu(p, \Theta_0(\hat{P}_n)) = \min_{w\in\mathbf{R}_+^{2d_F}} \{\langle w, F_S(\int m_S(x,\theta)d\hat{P}_n(x))\rangle \ \text{ s.t. } \ A'w = p\} \ , \tag{G.11}$$

where A and $v \mapsto F_S(v)$ are as defined in Example 2.1, and $m_S(x,\theta)$ is constant in $\theta \in \Theta$ (see (F.3)). Moreover, the minimizers of (G.11) are the Lagrange multipliers $\lambda(p, \hat{P}_n)$ of the primal problem that defines $\nu(p, \Theta_0(\hat{P}_n))$. Therefore, by (22) and direct calculation, solving (G.11) suffices for computing the bootstrap process G_n^* given by:

$$G_n^*(p) = -\lambda(p, \hat{P}_n)'\nabla F_S(\frac{1}{n}\sum_{i=1}^n m_S(X_i,\theta))\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\{m_S(X_i,\theta) - \frac{1}{n}\sum_{i=1}^n m_S(X_i,\theta)\} \ . \tag{G.12}$$

In our simulations we draw W_i from the Rademacher distribution – i.e. $P(W_i = 1) = P(W_i = -1) = 1/2$ – and we compute the critical value $\hat{c}_{1-\alpha}$ as the $1-\alpha$ quantile across bootstrap replications of:

$$\sup_{\gamma\in\mathcal{G}} \max\{G_n^*(p(\gamma)), 0\} \ , \tag{G.13}$$

where $p(\gamma)$ and $\mathcal{G}$ are as in (G.10). The support function for the confidence region $\hat{\Theta}_n^{\hat{c}_{1-\alpha}/\sqrt{n}}$ (as in Example 5.1) is then given by $\nu(\cdot, \hat{\Theta}_n) + \hat{c}_{1-\alpha}/\sqrt{n}$, and hence we check whether $\Theta_0(P) \subseteq \hat{\Theta}_n^{\hat{c}_{1-\alpha}/\sqrt{n}}$ by verifying that $\nu(p(\gamma), \Theta_0(P)) \leq \nu(p(\gamma), \hat{\Theta}_n) + \hat{c}_{1-\alpha}/\sqrt{n}$ for all $\gamma \in \mathcal{G}$ – see also Beresteanu and Molinari (2008). ■

Remark G.3. In order to compute $\hat{c}^B_{1-\alpha}(\tau_n)$ we draw samples $\{X_i^*\}_{i=1}^n$ from $\{X_i\}_{i=1}^n$ with replacement, let $\hat{P}_n^*$ denote the empirical measure induced by $\{X_i^*\}_{i=1}^n$ and $(\hat{\sigma}_n^{*(i)})^2$ be the corresponding estimate of the asymptotic variance of constraint number i. We then obtain $\hat{c}^B_{1-\alpha}(\tau_n)$ by computing the $1-\alpha$ quantile across bootstrap replications of:

$$\sup_{\theta\in\hat{\Theta}_n(\tau_n)} \max_{1\leq i\leq d_F} \Big\{\sqrt{n}(\frac{1}{\hat{\sigma}_n^{*(i)}}F^{(i)}(\int m(x,\theta)d\hat{P}_n^*(x)) - \frac{1}{\hat{\sigma}_n^{(i)}}F^{(i)}(\int m(x,\theta)d\hat{P}_n(x)))_+ \times \omega_n^{(i)}(\theta)\Big\} \ , \tag{G.14}$$

where $\omega_n^{(i)}(\theta) \equiv 1\{|F^{(i)}(\int m(x,\theta)d\hat{P}_n(x))| \leq \tau_n\hat{\sigma}_n^{(i)}/\sqrt{n}\}$. Since $CS(\tau_n)$ is a convex polygon, we compute its support function in a manner analogous to (G.9), and check whether $\Theta_0(P) \subseteq CS(\tau_n)$ by verifying that $\nu(p(\gamma), \Theta_0(P)) \leq \nu(p(\gamma), CS(\tau_n))$ for all $\gamma \in \mathcal{G}$, where $p(\gamma)$ and $\mathcal{G}$ are as in (G.10). ■

Remark G.4. Following the construction of $\hat{c}^B_{1-\alpha}(\tau_n)$, to obtain $\hat{c}^{AS}_{1-\alpha}(\theta)$ we draw samples $\{X_i^*\}_{i=1}^n$ from $\{X_i\}_{i=1}^n$ with replacement, let $\hat{P}_n^*$ denote the empirical measure induced by $\{X_i^*\}_{i=1}^n$ and $(\hat{\sigma}_n^{*(i)})^2$ be the corresponding estimate of the asymptotic variance of constraint i. For $\omega_n^{(i)}(\theta) \equiv 1\{|F^{(i)}(\int m(x,\theta)d\hat{P}_n(x))| \leq \tau_n\hat{\sigma}_n^{(i)}/\sqrt{n}\}$ and

$$Q_n^*(\theta) \equiv \max_{1\leq i\leq d_F} \Big\{(\frac{1}{\hat{\sigma}_n^{*(i)}}F^{(i)}(\int m(x,\theta)d\hat{P}_n^*(x)) - \frac{1}{\hat{\sigma}_n^{(i)}}F^{(i)}(\int m(x,\theta)d\hat{P}_n(x)))_+ \times \omega_n^{(i)}(\theta)\Big\} \ , \tag{G.15}$$

we then let $\hat{c}^{AS}_{1-\alpha}(\theta)$ be the $1-\alpha$ quantile of $\sqrt{n}Q_n^*(\theta)$ across 200 bootstrap replications. ■

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

Table 1: Proportion of simulated samples with empty set estimators.

	Estimator $\Theta_0(\hat{P}_n)$								
	$K = 5$			$K = 9$			$K = 15$		
Sample Size	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
$n = 200$	–	–	–	0.201	–	–	0.792	0.010	–
$n = 500$	–	–	–	0.035	–	–	0.420	–	–
$n = 1000$	–	–	–	0.003	–	–	0.152	–	–

	Estimator $\hat{\Theta}_n(\tau_n)$ with $\tau_n = \log(\log(n))$								
	$K = 5$			$K = 9$			$K = 15$		
Sample Size	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
$n = 200$	–	–	–	–	–	–	0.007	–	–
$n = 500$	–	–	–	–	–	–	–	–	–
$n = 1000$	–	–	–	–	–	–	–	–	–

Table 2: Median Hausdorff Distance.

	$n = 200$								
	$K = 5$			$K = 9$			$K = 15$		
Estimator	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Efficient	0.131	0.132	0.132	0.232	0.208	0.209	Inf	0.332	0.332
$\tau_n = \log(\log(n))$	0.372	0.372	0.372	0.423	0.423	0.423	0.393	0.392	0.392
$\tau_n = \log(n)$	0.941	0.941	0.941	1.138	1.138	1.138	1.226	1.226	1.226
$\tau_n = n^{1/8}$	0.414	0.414	0.414	0.476	0.476	0.476	0.455	0.455	0.455
$\tau_n = n^{1/4}$	0.702	0.702	0.702	0.838	0.838	0.838	0.879	0.879	0.879

	$n = 500$								
	$K = 5$			$K = 9$			$K = 15$		
Estimator	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Efficient	0.080	0.081	0.081	0.136	0.130	0.131	0.290	0.205	0.204
$\tau_n = \log(\log(n))$	0.251	0.251	0.251	0.316	0.316	0.316	0.315	0.315	0.315
$\tau_n = \log(n)$	0.692	0.692	0.692	0.890	0.890	0.890	1.021	1.021	1.021
$\tau_n = n^{1/8}$	0.285	0.285	0.285	0.362	0.362	0.362	0.371	0.371	0.371
$\tau_n = n^{1/4}$	0.542	0.542	0.542	0.696	0.696	0.696	0.783	0.783	0.783

	$n = 1000$								
	$K = 5$			$K = 9$			$K = 15$		
Estimator	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Efficient	0.058	0.058	0.058	0.093	0.092	0.093	0.172	0.144	0.144
$\tau_n = \log(\log(n))$	0.185	0.185	0.185	0.244	0.244	0.244	0.257	0.257	0.257
$\tau_n = \log(n)$	0.537	0.537	0.537	0.713	0.713	0.713	0.841	0.841	0.841
$\tau_n = n^{1/8}$	0.216	0.216	0.216	0.285	0.285	0.285	0.308	0.308	0.308
$\tau_n = n^{1/4}$	0.447	0.447	0.447	0.592	0.592	0.592	0.690	0.690	0.690

Table 3: Set Confidence Region Coverage Probability. Nominal Coverage = 0.95.

	$n = 200$								
	$K = 5$			$K = 9$			$K = 15$		
Procedure	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Efficient	0.945	0.942	0.940	0.790	0.913	0.895	0.208	0.885	0.820
B. $\tau_n = \log(\log(n))$	0.980	0.984	0.986	0.990	0.992	0.992	0.989	0.997	0.998
B. $\tau_n = \log(n)$	0.998	0.998	0.998	1.000	1.000	1.000	1.000	1.000	1.000
B. $\tau_n = n^{1/8}$	0.983	0.986	0.987	0.993	0.993	0.994	0.994	0.998	0.999
B. $\tau_n = n^{1/4}$	0.994	0.995	0.995	0.999	0.999	0.999	1.000	1.000	1.000
	$n = 500$								
	$K = 5$			$K = 9$			$K = 15$		
Procedure	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Efficient	0.950	0.940	0.940	0.946	0.916	0.916	0.573	0.879	0.870
B. $\tau_n = \log(\log(n))$	0.967	0.972	0.978	0.983	0.982	0.982	0.986	0.988	0.990
B. $\tau_n = \log(n)$	0.994	0.993	0.995	0.999	0.999	0.998	1.000	1.000	1.000
B. $\tau_n = n^{1/8}$	0.971	0.975	0.979	0.987	0.984	0.986	0.990	0.991	0.991
B. $\tau_n = n^{1/4}$	0.989	0.989	0.990	0.998	0.998	0.997	0.999	1.000	1.000
	$n = 1000$								
	$K = 5$			$K = 9$			$K = 15$		
Procedure	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Efficient	0.959	0.946	0.946	0.975	0.926	0.925	0.829	0.891	0.889
B. $\tau_n = \log(\log(n))$	0.969	0.971	0.979	0.983	0.981	0.980	0.981	0.983	0.982
B. $\tau_n = \log(n)$	0.991	0.993	0.994	0.998	0.998	0.997	1.000	1.000	0.999
B. $\tau_n = n^{1/8}$	0.970	0.973	0.981	0.987	0.984	0.983	0.985	0.986	0.986
B. $\tau_n = n^{1/4}$	0.989	0.989	0.992	0.997	0.996	0.994	0.999	0.999	0.998

Table 4: Set Confidence Region Median Hausdorff Distance. Nominal Coverage 0.95.

	$n = 200$								
	$K = 5$			$K = 9$			$K = 15$		
Procedure	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Efficient	0.439	0.430	0.430	0.639	0.502	0.498	Inf	0.542	0.524
B. $\tau_n = \log(\log(n))$	0.566	0.576	0.585	0.826	0.855	0.871	1.504	1.645	1.724
B. $\tau_n = \log(n)$	0.710	0.718	0.727	1.245	1.259	1.275	3.257	3.359	3.413
B. $\tau_n = n^{1/8}$	0.577	0.585	0.594	0.855	0.882	0.897	1.629	1.764	1.834
B. $\tau_n = n^{1/4}$	0.645	0.651	0.660	1.058	1.075	1.092	2.505	2.593	2.672

	$n = 500$								
	$K = 5$			$K = 9$			$K = 15$		
Procedure	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Efficient	0.276	0.273	0.273	0.398	0.349	0.349	0.740	0.371	0.373
B. $\tau_n = \log(\log(n))$	0.328	0.334	0.344	0.468	0.481	0.488	0.580	0.600	0.609
B. $\tau_n = \log(n)$	0.384	0.387	0.392	0.597	0.603	0.608	0.862	0.876	0.883
B. $\tau_n = n^{1/8}$	0.332	0.339	0.347	0.478	0.490	0.496	0.600	0.618	0.626
B. $\tau_n = n^{1/4}$	0.366	0.369	0.374	0.551	0.557	0.562	0.759	0.773	0.779

	$n = 1000$								
	$K = 5$			$K = 9$			$K = 15$		
Procedure	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Efficient	0.195	0.193	0.193	0.282	0.262	0.262	0.388	0.282	0.283
B. $\tau_n = \log(\log(n))$	0.226	0.227	0.237	0.326	0.335	0.341	0.394	0.405	0.411
B. $\tau_n = \log(n)$	0.257	0.258	0.261	0.389	0.392	0.394	0.514	0.519	0.523
B. $\tau_n = n^{1/8}$	0.228	0.230	0.239	0.333	0.341	0.345	0.404	0.414	0.419
B. $\tau_n = n^{1/4}$	0.250	0.252	0.254	0.372	0.376	0.378	0.480	0.485	0.489

Table 5: Empirical Size $H_0 : \theta_F \in \Theta_0(P)$ (on Flat Face). Nominal Size = 0.05.

	$n=200$								
	$K=5$			$K=9$			$K=15$		
Procedure	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
Eff. $\kappa_n=\log(\log(n))$	0.037	0.055	0.056	0.205	0.066	0.073	0.792	0.113	0.146
Eff. $\kappa_n=\log(n)$	0.034	0.054	0.056	0.204	0.057	0.067	0.792	0.089	0.134
Eff. $\kappa_n=n^{1/8}$	0.036	0.054	0.056	0.205	0.065	0.072	0.792	0.110	0.144
Eff. $\kappa_n=n^{1/4}$	0.035	0.054	0.056	0.204	0.058	0.068	0.792	0.094	0.136
A.S. $\tau_n=\log(\log(n))$	0.040	0.040	0.039	0.012	0.016	0.015	0.004	0.006	0.007
A.S. $\tau_n=\log(n)$	0.011	0.017	0.019	0.006	0.008	0.009	0.003	0.004	0.004
A.S. $\tau_n=n^{1/8}$	0.039	0.039	0.039	0.012	0.014	0.014	0.003	0.006	0.007
A.S. $\tau_n=n^{1/4}$	0.018	0.024	0.026	0.007	0.011	0.011	0.003	0.004	0.005

	$n=500$								
	$K=5$			$K=9$			$K=15$		
Procedure	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
Eff. $\kappa_n=\log(\log(n))$	0.040	0.052	0.052	0.045	0.053	0.053	0.421	0.090	0.093
Eff. $\kappa_n=\log(n)$	0.034	0.052	0.052	0.040	0.047	0.047	0.420	0.076	0.082
Eff. $\kappa_n=n^{1/8}$	0.039	0.052	0.052	0.044	0.052	0.052	0.420	0.089	0.092
Eff. $\kappa_n=n^{1/4}$	0.035	0.052	0.052	0.040	0.048	0.048	0.420	0.079	0.084
A.S. $\tau_n=\log(\log(n))$	0.049	0.050	0.049	0.017	0.022	0.022	0.012	0.018	0.017
A.S. $\tau_n=\log(n)$	0.016	0.027	0.027	0.007	0.012	0.011	0.006	0.008	0.010
A.S. $\tau_n=n^{1/8}$	0.049	0.050	0.049	0.016	0.021	0.021	0.011	0.016	0.016
A.S. $\tau_n=n^{1/4}$	0.024	0.043	0.041	0.008	0.014	0.013	0.008	0.011	0.012

	$n=1000$								
	$K=5$			$K=9$			$K=15$		
Procedure	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$	$C=0.1$	$C=0.5$	$C=1$
Eff. $\kappa_n=\log(\log(n))$	0.048	0.049	0.049	0.054	0.056	0.056	0.189	0.105	0.105
Eff. $\kappa_n=\log(n)$	0.038	0.049	0.049	0.020	0.054	0.054	0.154	0.082	0.083
Eff. $\kappa_n=n^{1/8}$	0.030	0.048	0.048	0.011	0.048	0.048	0.152	0.070	0.071
Eff. $\kappa_n=n^{1/4}$	0.037	0.048	0.048	0.017	0.054	0.054	0.154	0.080	0.081
A.S. $\tau_n=\log(\log(n))$	0.050	0.050	0.048	0.026	0.027	0.028	0.016	0.020	0.051
A.S. $\tau_n=\log(n)$	0.023	0.045	0.043	0.008	0.011	0.011	0.006	0.009	0.020
A.S. $\tau_n=n^{1/8}$	0.050	0.050	0.048	0.020	0.023	0.024	0.014	0.018	0.009
A.S. $\tau_n=n^{1/4}$	0.024	0.049	0.048	0.008	0.014	0.015	0.007	0.011	0.017

Table 6: Empirical Size $H_0 : \theta_K \in \Theta_0(P)$ (on Kink). Nominal Size = 0.05.

	$n = 200$								
	$K = 5$			$K = 9$			$K = 15$		
Procedure	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Eff. $\kappa_n = \log(\log(n))$	0.028	0.044	0.062	0.214	0.056	0.090	0.796	0.110	0.149
Eff. $\kappa_n = \log(n)$	0.024	0.005	0.024	0.206	0.017	0.017	0.793	0.046	0.046
Eff. $\kappa_n = n^{1/8}$	0.028	0.039	0.057	0.213	0.050	0.080	0.795	0.098	0.137
Eff. $\kappa_n = n^{1/4}$	0.025	0.013	0.037	0.208	0.025	0.035	0.793	0.060	0.073
A.S. $\tau_n = \log(\log(n))$	0.028	0.035	0.027	0.016	0.013	0.017	0.003	0.003	0.005
A.S. $\tau_n = \log(n)$	0.020	0.019	0.023	0.010	0.010	0.013	0.002	0.002	0.004
A.S. $\tau_n = n^{1/8}$	0.027	0.032	0.026	0.015	0.012	0.015	0.003	0.002	0.005
A.S. $\tau_n = n^{1/4}$	0.023	0.020	0.025	0.012	0.010	0.013	0.003	0.002	0.004

	$n = 500$								
	$K = 5$			$K = 9$			$K = 15$		
Procedure	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Eff. $\kappa_n = \log(\log(n))$	0.010	0.048	0.055	0.060	0.047	0.087	0.435	0.089	0.119
Eff. $\kappa_n = \log(n)$	0.009	0.007	0.029	0.045	0.012	0.017	0.426	0.024	0.029
Eff. $\kappa_n = n^{1/8}$	0.010	0.043	0.052	0.058	0.040	0.078	0.434	0.079	0.106
Eff. $\kappa_n = n^{1/4}$	0.009	0.016	0.037	0.048	0.015	0.033	0.428	0.034	0.047
A.S. $\tau_n = \log(\log(n))$	0.026	0.045	0.029	0.023	0.020	0.028	0.018	0.016	0.017
A.S. $\tau_n = \log(n)$	0.020	0.017	0.024	0.013	0.010	0.019	0.011	0.011	0.013
A.S. $\tau_n = n^{1/8}$	0.026	0.044	0.028	0.022	0.019	0.026	0.017	0.015	0.017
A.S. $\tau_n = n^{1/4}$	0.023	0.023	0.024	0.016	0.011	0.020	0.012	0.012	0.015

	$n = 1000$								
	$K = 5$			$K = 9$			$K = 15$		
Procedure	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$	$C = 0.1$	$C = 0.5$	$C = 1$
Eff. $\kappa_n = \log(\log(n))$	0.006	0.050	0.054	0.037	0.047	0.082	0.197	0.089	0.107
Eff. $\kappa_n = \log(n)$	0.002	0.012	0.033	0.016	0.007	0.022	0.175	0.023	0.027
Eff. $\kappa_n = n^{1/8}$	0.004	0.044	0.050	0.033	0.037	0.073	0.194	0.075	0.096
Eff. $\kappa_n = n^{1/4}$	0.002	0.020	0.037	0.018	0.010	0.033	0.178	0.032	0.041
A.S. $\tau_n = \log(\log(n))$	0.029	0.053	0.038	0.024	0.024	0.030	0.025	0.017	0.061
A.S. $\tau_n = \log(n)$	0.022	0.020	0.023	0.013	0.011	0.018	0.012	0.010	0.024
A.S. $\tau_n = n^{1/8}$	0.026	0.052	0.033	0.022	0.021	0.028	0.022	0.015	0.016
A.S. $\tau_n = n^{1/4}$	0.022	0.035	0.023	0.016	0.012	0.019	0.014	0.011	0.022

Table 7: Empirical Power $H_0 : \theta_F \in \Theta_0(P)$ (on Flat Face). Nominal Size = 0.05.

	$n = 200$ and $K = 5$									
	$C = 0.5$					$C = 1$				
Procedure	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$
Eff. $\kappa_n = \log(\log(n))$	0.055	0.306	0.722	0.951	0.996	0.056	0.306	0.722	0.951	0.996
Eff. $\kappa_n = \log(n)$	0.054	0.306	0.722	0.951	0.996	0.056	0.306	0.722	0.951	0.996
Eff. $\kappa_n = n^{1/8}$	0.054	0.306	0.722	0.951	0.996	0.056	0.306	0.722	0.951	0.996
Eff. $\kappa_n = n^{1/4}$	0.054	0.306	0.722	0.951	0.996	0.056	0.306	0.722	0.951	0.996
A.S. $\tau_n = \log(\log(n))$	0.040	0.227	0.550	0.837	0.971	0.039	0.231	0.550	0.843	0.970
A.S. $\tau_n = \log(n)$	0.017	0.144	0.483	0.827	0.970	0.019	0.143	0.484	0.830	0.969
A.S. $\tau_n = n^{1/8}$	0.039	0.219	0.536	0.833	0.971	0.039	0.221	0.536	0.839	0.970
A.S. $\tau_n = n^{1/4}$	0.024	0.155	0.488	0.829	0.971	0.026	0.158	0.489	0.833	0.970

	$n = 200$ and $K = 9$									
	$C = 0.5$					$C = 1$				
Procedure	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$
Eff. $\kappa_n = \log(\log(n))$	0.066	0.293	0.685	0.943	0.996	0.073	0.295	0.686	0.943	0.996
Eff. $\kappa_n = \log(n)$	0.057	0.271	0.672	0.940	0.996	0.067	0.279	0.674	0.941	0.996
Eff. $\kappa_n = n^{1/8}$	0.065	0.290	0.682	0.943	0.996	0.072	0.293	0.683	0.943	0.996
Eff. $\kappa_n = n^{1/4}$	0.058	0.276	0.674	0.941	0.996	0.068	0.282	0.675	0.941	0.996
A.S. $\tau_n = \log(\log(n))$	0.016	0.074	0.225	0.488	0.744	0.015	0.072	0.232	0.494	0.745
A.S. $\tau_n = \log(n)$	0.008	0.045	0.186	0.447	0.728	0.009	0.049	0.190	0.455	0.727
A.S. $\tau_n = n^{1/8}$	0.014	0.071	0.222	0.486	0.743	0.014	0.068	0.229	0.491	0.744
A.S. $\tau_n = n^{1/4}$	0.011	0.059	0.207	0.470	0.735	0.011	0.058	0.209	0.473	0.736

	$n = 1000$ and $K = 5$									
	$C = 0.5$					$C = 1$				
Procedure	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$
Eff. $\kappa_n = \log(\log(n))$	0.049	0.285	0.702	0.954	0.998	0.049	0.285	0.702	0.954	0.998
Eff. $\kappa_n = \log(n)$	0.048	0.285	0.701	0.954	0.998	0.048	0.285	0.701	0.954	0.998
Eff. $\kappa_n = n^{1/8}$	0.048	0.285	0.702	0.954	0.998	0.048	0.285	0.702	0.954	0.998
Eff. $\kappa_n = n^{1/4}$	0.048	0.285	0.701	0.954	0.998	0.048	0.285	0.701	0.954	0.998
A.S. $\tau_n = \log(\log(n))$	0.050	0.295	0.709	0.952	0.998	0.048	0.294	0.708	0.954	0.997
A.S. $\tau_n = \log(n)$	0.045	0.223	0.566	0.884	0.988	0.043	0.221	0.567	0.884	0.988
A.S. $\tau_n = n^{1/8}$	0.050	0.295	0.709	0.952	0.997	0.048	0.294	0.708	0.954	0.997
A.S. $\tau_n = n^{1/4}$	0.049	0.282	0.645	0.903	0.988	0.048	0.282	0.646	0.904	0.988

	$n = 1000$ and $K = 9$									
	$C = 0.5$					$C = 1$				
Procedure	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$
Eff. $\kappa_n = \log(\log(n))$	0.054	0.209	0.529	0.851	0.987	0.054	0.209	0.529	0.851	0.987
Eff. $\kappa_n = \log(n)$	0.048	0.193	0.508	0.844	0.985	0.048	0.194	0.508	0.844	0.985
Eff. $\kappa_n = n^{1/8}$	0.054	0.208	0.526	0.850	0.987	0.054	0.208	0.526	0.850	0.987
Eff. $\kappa_n = n^{1/4}$	0.050	0.197	0.509	0.844	0.985	0.050	0.198	0.509	0.844	0.985
A.S. $\tau_n = \log(\log(n))$	0.027	0.109	0.333	0.679	0.926	0.028	0.112	0.332	0.680	0.927
A.S. $\tau_n = \log(n)$	0.011	0.072	0.256	0.600	0.894	0.011	0.071	0.255	0.595	0.892
A.S. $\tau_n = n^{1/8}$	0.023	0.106	0.330	0.676	0.921	0.024	0.107	0.329	0.675	0.922
A.S. $\tau_n = n^{1/4}$	0.014	0.081	0.269	0.604	0.895	0.015	0.082	0.269	0.601	0.894

Table 8: Empirical Power $H_0 : \theta_K \in \Theta_0(P)$ (on Kink). Nominal Size = 0.05.

	$n = 200$ and $K = 5$									
	$C = 0.5$					$C = 1$				
Procedure	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$
Eff. $\kappa_n = \log(\log(n))$	0.044	0.904	1.000	1.000	1.000	0.062	0.977	1.000	1.000	1.000
Eff. $\kappa_n = \log(n)$	0.005	0.526	0.998	1.000	1.000	0.024	0.921	1.000	1.000	1.000
Eff. $\kappa_n = n^{1/8}$	0.039	0.892	1.000	1.000	1.000	0.057	0.976	1.000	1.000	1.000
Eff. $\kappa_n = n^{1/4}$	0.013	0.734	0.999	1.000	1.000	0.037	0.957	1.000	1.000	1.000
A.S. $\tau_n = \log(\log(n))$	0.035	0.784	1.000	1.000	1.000	0.027	0.896	1.000	1.000	1.000
A.S. $\tau_n = \log(n)$	0.019	0.751	1.000	1.000	1.000	0.023	0.891	1.000	1.000	1.000
A.S. $\tau_n = n^{1/8}$	0.032	0.781	1.000	1.000	1.000	0.026	0.896	1.000	1.000	1.000
A.S. $\tau_n = n^{1/4}$	0.020	0.762	1.000	1.000	1.000	0.025	0.896	1.000	1.000	1.000

	$n = 200$ and $K = 9$									
	$C = 0.5$					$C = 1$				
Procedure	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$
Eff. $\kappa_n = \log(\log(n))$	0.056	0.665	0.986	1.000	1.000	0.090	0.895	1.000	1.000	1.000
Eff. $\kappa_n = \log(n)$	0.017	0.346	0.963	0.999	1.000	0.017	0.577	0.995	1.000	1.000
Eff. $\kappa_n = n^{1/8}$	0.050	0.632	0.985	1.000	1.000	0.080	0.872	0.999	1.000	1.000
Eff. $\kappa_n = n^{1/4}$	0.025	0.457	0.976	0.999	1.000	0.035	0.711	0.997	1.000	1.000
A.S. $\tau_n = \log(\log(n))$	0.013	0.322	0.881	0.970	0.983	0.017	0.495	0.939	0.979	0.987
A.S. $\tau_n = \log(n)$	0.010	0.313	0.881	0.970	0.983	0.013	0.481	0.939	0.979	0.987
A.S. $\tau_n = n^{1/8}$	0.012	0.321	0.881	0.970	0.983	0.015	0.494	0.939	0.979	0.987
A.S. $\tau_n = n^{1/4}$	0.010	0.315	0.881	0.970	0.983	0.013	0.490	0.939	0.979	0.987

	$n = 1000$ and $K = 5$									
	$C = 0.5$					$C = 1$				
Procedure	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$
Eff. $\kappa_n = \log(\log(n))$	0.050	0.937	1.000	1.000	1.000	0.054	0.961	1.000	1.000	1.000
Eff. $\kappa_n = \log(n)$	0.012	0.811	1.000	1.000	1.000	0.033	0.931	1.000	1.000	1.000
Eff. $\kappa_n = n^{1/8}$	0.044	0.934	1.000	1.000	1.000	0.050	0.960	1.000	1.000	1.000
Eff. $\kappa_n = n^{1/4}$	0.020	0.864	1.000	1.000	1.000	0.037	0.944	1.000	1.000	1.000
A.S. $\tau_n = \log(\log(n))$	0.053	0.917	1.000	1.000	1.000	0.038	0.899	1.000	1.000	1.000
A.S. $\tau_n = \log(n)$	0.020	0.848	1.000	1.000	1.000	0.023	0.899	1.000	1.000	1.000
A.S. $\tau_n = n^{1/8}$	0.052	0.908	1.000	1.000	1.000	0.033	0.899	1.000	1.000	1.000
A.S. $\tau_n = n^{1/4}$	0.035	0.869	1.000	1.000	1.000	0.023	0.899	1.000	1.000	1.000

	$n = 1000$ and $K = 9$									
	$C = 0.5$					$C = 1$				
Procedure	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$	$h = 0$	$h = 2.5$	$h = 5$	$h = 7.5$	$h = 10$
Eff. $\kappa_n = \log(\log(n))$	0.047	0.601	0.995	1.000	1.000	0.082	0.944	1.000	1.000	1.000
Eff. $\kappa_n = \log(n)$	0.007	0.303	0.979	1.000	1.000	0.022	0.661	1.000	1.000	1.000
Eff. $\kappa_n = n^{1/8}$	0.037	0.547	0.993	1.000	1.000	0.073	0.935	1.000	1.000	1.000
Eff. $\kappa_n = n^{1/4}$	0.010	0.331	0.983	1.000	1.000	0.033	0.780	1.000	1.000	1.000
A.S. $\tau_n = \log(\log(n))$	0.024	0.532	0.999	1.000	1.000	0.030	0.829	1.000	1.000	1.000
A.S. $\tau_n = \log(n)$	0.011	0.473	0.999	1.000	1.000	0.018	0.803	1.000	1.000	1.000
A.S. $\tau_n = n^{1/8}$	0.021	0.524	0.999	1.000	1.000	0.028	0.823	1.000	1.000	1.000
A.S. $\tau_n = n^{1/4}$	0.012	0.486	0.999	1.000	1.000	0.019	0.803	1.000	1.000	1.000

Table 9: Median Confidence Region Computation Time in Seconds.

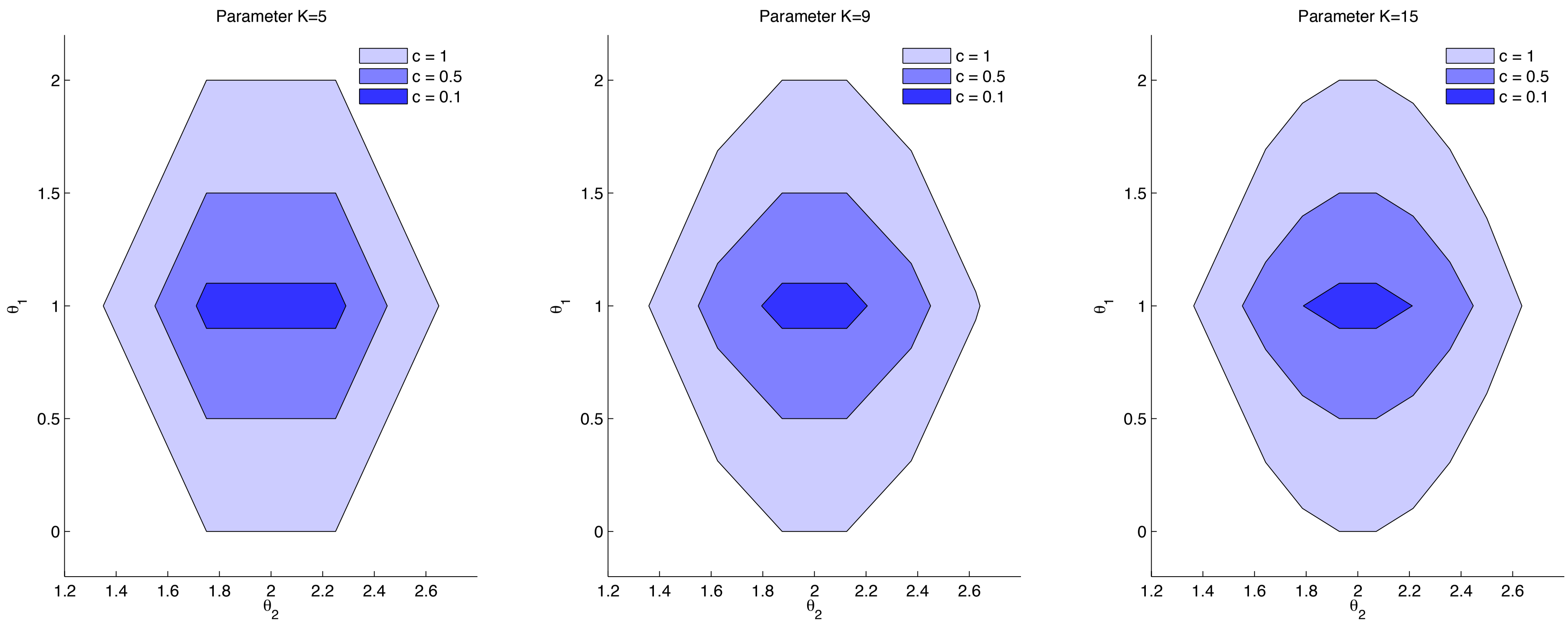
Procedure	$K=5$, $C=0.1$	$K=5$, $C=0.5$	$K=5$, $C=1$	$K=9$, $C=0.1$	$K=9$, $C=0.5$	$K=9$, $C=1$	$K=15$, $C=0.1$	$K=15$, $C=0.5$	$K=15$, $C=1$
$n=200$									
Efficient	2.528	2.683	2.751	3.741	4.061	4.319	5.824	6.138	6.514
B. $\tau_n=\log(\log(n))$	1.997	1.946	1.917	2.308	2.436	2.535	3.023	3.342	3.501
B. $\tau_n=\log(n)$	1.925	1.890	1.907	2.434	2.492	2.545	3.311	3.431	3.495
B. $\tau_n=n^{1/8}$	1.995	1.944	1.917	2.335	2.455	2.546	3.086	3.368	3.519
B. $\tau_n=n^{1/4}$	1.976	1.923	1.921	2.421	2.503	2.572	3.281	3.452	3.542
$n=500$									
Efficient	2.577	2.691	2.730	3.805	4.191	4.493	5.867	6.397	6.839
B. $\tau_n=\log(\log(n))$	2.086	2.002	1.936	2.420	2.554	2.660	3.277	3.538	3.741
B. $\tau_n=\log(n)$	2.007	1.947	1.919	2.543	2.607	2.678	3.536	3.627	3.749
B. $\tau_n=n^{1/8}$	2.082	1.998	1.936	2.442	2.565	2.670	3.344	3.565	3.758
B. $\tau_n=n^{1/4}$	2.049	1.983	1.933	2.534	2.617	2.702	3.528	3.644	3.797
$n=1000$									
Efficient	2.587	2.649	2.654	3.758	4.203	4.484	5.742	6.383	6.952
B. $\tau_n=\log(\log(n))$	2.174	2.055	1.979	2.532	2.656	2.754	3.366	3.626	3.767
B. $\tau_n=\log(n)$	2.086	2.010	1.947	2.641	2.709	2.786	3.613	3.729	3.802
B. $\tau_n=n^{1/8}$	2.166	2.049	1.980	2.551	2.673	2.769	3.427	3.668	3.804
B. $\tau_n=n^{1/4}$	2.124	2.034	1.963	2.653	2.728	2.812	3.613	3.760	3.854

Table 10: Statistics Conditional on $\Theta_0(\hat{P}_n)\neq\emptyset$.

Specification	Med. $d_H(\Theta_0(\hat{P}_n),\Theta_0(P))$	$\Theta_0(P)$ CI Coverage	θ_0 on Flat Face Size	θ_0 on Kink Size
$n=200$				
$K=9$, $C=0.1$	0.200	0.989	0.005	0.016
$K=15$, $C=0.1$	0.250	0.998	0.000	0.017
$n=500$				
$K=9$, $C=0.1$	0.133	0.980	0.010	0.026
$K=15$, $C=0.1$	0.202	0.987	0.002	0.027
$n=200$				
$K=9$, $C=0.1$	0.093	0.978	0.017	0.034
$K=15$, $C=0.1$	0.157	0.978	0.003	0.054

Note: Empirical size for tests of $H_0:\theta_0\in\Theta_0(P)$ reported for $\kappa_n=\log(\log(n))$.

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



References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

Andrews, D. W. K. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, **78** 119–157.

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. (2012). Set identified linear models. *Econometrica*, **80** 1129–1155.

Bugni, F. A. (2010). Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica*, **78** 735–753.

Chernozhukov, V., Hong, H. and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, **75** 1243–1284.

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998). *Local Properties of Distribuions of Stochastic Functionals*. American Mathematical Society, Providence.

Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Springer, New York.

Ho, K. (2009). Insurer-provider networks in the medical care market. *American Economic Review*, **99** 1039–1079.

Kaido, H. (2010). A dual approach to inference for partially identified econometric models. Working Paper. Boston University.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag, New York.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, 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.

Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, **70** 583–601.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton.

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.